R E P O R T   R E S U M E S

ED 016 535   24   PS 000 451

THE PERFORMANCE OF FIRST GRADE CHILDREN IN FOUR LEVELS OF CONSERVATION OF NUMEROUSNESS AND THREE IQ GROUPS WHEN SOLVING ARITHMETIC ADDITION PROBLEMS.

BY- STEFFE, LESLIE P.
WISCONSIN UNIV., MADISON

REPORT NUMBER UW-R/D-CENT-TR-14   PUB DATE DEC 66
REPORT NUMBER BR-5-0216-TR-14
CONTRACT OEC-5-10-154
EDRS PRICE MF-$0.50 HC-$2.76   67P.

DESCRIPTORS- ANALYSIS OF VARIANCE, *GRADE 1, ARITHMETIC, *INTELLIGENCE QUOTIENT, CONCEPT FORMATION, *NUMBER CONCEPTS, *PROBLEM SOLVING, *PERFORMANCE, CORRELATION STUDIES, STIMULUS DEVICES, PIAGET, KUHLMAN ANDERSON INTELLIGENCE TEST, RACINE

ACCORDING TO PIAGET, THE CONCEPT OF CONSERVATION IS A PREREQUISITE TO MATHEMATICAL UNDERSTANDING. THIS STUDY SOUGHT TO DETERMINE WHETHER THE ABILITY OF FIRST GRADE CHILDREN TO SOLVE ADDITION PROBLEMS WAS DEPENDENT ON THEIR MASTERY OF CONSERVATION OF NUMEROUSNESS. SECONDARY PURPOSES WERE TO INVESTIGATE THE EFFECT IN PROBLEM STATEMENTS OF (1) PHYSICAL OR PICTORIAL AIDS, AND (2) THE PRESENCE OR ABSENCE OF TRANSFORMATIONS. A TEST OF CONSERVATION OF NUMEROUSNESS THAT DIVIDED CHILDREN INTO 4 LEVELS WAS DEVELOPED IN A PILOT STUDY. THE POPULATION FOR THE MAIN STUDY WAS 2,166 FIRST GRADE CHILDREN WHO HAD PROGRESSED THROUGH ABOUT 3/4 OF AN ARITHMETIC CURRICULUM. ALL WERE GIVEN AN IQ TEST, WHICH WAS USED TO DEFINE 3 IQ LEVELS, AND THEN 341 CHILDREN WERE RANDOMLY SELECTED AND GIVEN (1) THE CONSERVATION OF NUMEROUSNESS TEST, (2) A TEST WITH 18 ADDITION PROBLEMS, EACH HAVING EITHER (A) PHYSICAL, (B) PICTORIAL, OR (C) NO AIDS, AND EITHER HAVING OR LACKING A TRANSFORMATION, AND (3) A TEST OF ADDITION FACTS. AN ANALYSIS OF VARIANCE EVALUATIONS BASED ON A SUBSAMPLE OF 121 CHILDREN INDICATED THAT BOTH THOSE AT THE LOWEST LEVEL OF CONSERVATION MASTERY AND THOSE AT THE LOWEST IQ LEVEL PERFORMED SIGNIFICANTLY LESS WELL ON TESTS (2) AND (3) ABOVE. FOR THE PROBLEM-SOLVING TEST, PROBLEMS HAVING A TRANSFORMATION WERE SIGNIFICANTLY EASIER THAN THOSE WITHOUT, AND PROBLEMS HAVING NO AIDS WERE SIGNIFICANTLY HARDER THAN THOSE WITH PHYSICAL OR PICTORIAL AIDS. CORRELATION OF .49 WAS FOUND BETWEEN THE PROBLEM-SOLVING AND ADDITION FACTS TESTS, WHICH WAS FELT TO INDICATE THAT ACTUAL PROBLEM-SOLVING RATHER THAN SIMPLE DRILL WAS NECESSARY TO LEARNING ADDITION FACTS. (DR)

BR 5-0216
PA 24

THE PERFORMANCE OF FIRST GRADE CHILDREN IN FOUR LEVELS OF CONSERVATION OF NUMEROUSNESS AND THREE IQ GROUPS WHEN SOLVING ARITHMETIC ADDITION PROBLEMS

RESEARCH CLR DEVELOPMENT

RESEARCH AND DEVELOPMENT
CENTER FOR LEARNING
AND RE-EDUCATION

Technical Report No. 14

THE PERFORMANCE OF FIRST GRADE CHILDREN IN FOUR LEVELS OF CONSERVATION OF NUMEROUSNESS AND THREE IQ GROUPS WHEN SOLVING ARITHMETIC ADDITION PROBLEMS

Leslie P. Steffe

Based on a doctoral dissertation under the direction of
Henry Van Engen, Professor of Education and Mathematics

Research and Development Center
for Learning and Re-education
The University of Wisconsin
Madison, Wisconsin
December 1966

The research reported herein was perfomed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare, under the provisions of the Cooperative Research Program

Center No. C–03 / Contract OE 5–10–154

OTHER REPORTS OF THE R & D CENTER FOR LEARNING AND RE-EDUCATION

TECHNICAL REPORTS

No. 1 Klausmeier, H. J., Davis, J. K., Ramsay, J. G., Fredrick, W. C., and Davies, Mary H. Concept learning and problem solving: A bibliography, 1950 - 1964. October 1965.

No. 2 Goodwin, W. L. The effects on achievement test results of varying conditions of experimental atmosphere, notice of test, test administration, and test scoring. November 1965.

No. 3 Fredrick, W. C. The effects of instructions, concept complexity, method of presentation, and order of concepts upon a concept attainment task. November 1965.

No. 4 Ramsay, J. G. The attainment of concepts from figural and verbal instances, by individuals and pairs. January 1966.

No. 5 Van Engen, H., and Steffe, L. P. First grade children's concept of addition of natural numbers. February 1966.

No. 6 Lynch, D. O. Concept identification as a function of instructions, labels, sequence, concept type, and test item type. June 1966.

No. 7 Biaggio, Angela M. B. Relative predictability of freshman grade-point averages from SAT scores in Negro and white Southern colleges. September 1966.

No. 8 Kalish, Patricia W. Concept attainment as a function of monetary incentives, competition, and instructions. September 1966.

No. 9 Baldwin, Thelma L., and Johnson, T. J. Teacher behaviors and effectiveness of reinforcement. September 1966.

No. 10 Fang, M. C. S. Effect of incentive and complexity on performance of students from two social class backgrounds on a concept identification task. September 1966.

No. 11 Lemke, E. A., Klausmeier, H. J., and Harris, C. W. The relationship of selected cognitive abilities to concept attainment and information processing. October 1966.

No. 12 Burmeister, Lou E. An evaluation of the inductive and deductive group approaches to teaching selected word analysis generalizations to disabled readers in eighth and ninth grade. November 1966.

No. 13 Boe, Barbara L. The ability of secondary school pupils to perceive the plane sections of selected solid figures. November 1966.

OCCASIONAL PAPERS

No. 1 Staats, A. W. Emotions and images in language: A learning analysis of their acquisition and function. June 1966.

No. 2 Davis, G. A. The current status of research and theory in human problem solving. June 1966.

No. 3 Klausmeier, H. J., Goodwin, W. L., Prasch, J., and Goodson, M. R. Project MODELS: Maximizing opportunities for development and experimentation in learning in the schools. July 1966.

No. 4 Otto, W. The relationship of reactive inhibition and school achievement: Theory, research, and implications. September 1966.

No. 5 Baker, F. B. The development of a computer model of the concept-attainment process. October 1966.

PREFACE

This report is based on the doctoral dissertation of Leslie P. Steffe. Members of the examining committee were Henry Van Engen, Chairman; Frank B. Baker; Milton A. Beckman; Eric Immel; and Wayne Otto.

The goal of the R & D Center for Learning and Re-education is the improvement of cognitive learning in children and adults, commensurate with good personality development. Activities are focused on three main problem areas: developing exemplary instructional systems; refining the science of human behavior and learning as well as the technology of instruction; and inventing new models for school experimentation, development activities, and so on. Through synthesizing present knowledge and conducting research to generate new knowledge, we are extending the understanding of human learning and the variables associated with efficiency of school learning.

This study is part of a program for the improvement of instruction in elementary mathematics under the direction of Professor Henry Van Engen. It illustrates the role of research in conjunction with the development of an exemplary instructional program by means of television and related material. Mr. Steffe found that some first grade children have not acquired the knowledge essential for learning addition and that it is possible to identify those children when they enter first grade. Two lines of further research are suggested: refining the instrument to predict relative success in the standard curriculum and identifying suitable experiences for those children who are not now profiting from the curriculum.

Herbert J. Klausmeier
Co-Director for Research

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

A test of conservation of numerousness was developed to separate first grade children into four different levels to study their relative performance when solving arithmetic addition problems. The children were also placed into three different IQ groups (78—100, 101—113, and 114—140) within each level, so that it was possible to study the relative performances of the children in the three IQ groups as well as the four levels when solving arithmetic addition problems. Two variables in problem solving were of interest: 1. The presence of a described transformation versus no described transformation of the sets in an addition problem. 2. The presence of (a) physical aids, (b) pictorial aids, and (c) no aids when solving addition problems. A study of the performances of the children in the four levels and three IQ groups on a test of addition facts was also conducted, as well as a correlation study of the scores of the children when solving addition problems and their scores when responding to addition facts.

The children who were in the lowest level of conservation of numerousness performed significantly less well than did the children who were in the upper three levels, who did not differ significantly when solving addition problems. The children who were in the lowest IQ group performed significantly less well than did the children who were in the upper two IQ groups, who did not differ significantly when solving addition problems.

The performance of the children in the four levels on the addition facts test was significantly different with the mean of the scores of the children in the fourth level the lowest. The performance of the children in the three IQ groups on the addition facts test was significantly different with the mean of the scores of the children in the IQ group 78-100 the lowest. The problems with a described transformation were significantly easier than the problems with no described transformation. The problems with no accompanying aids were significantly more difficult than the problems with either physical aids or pictorial aids, which did not differ.

A correlation of .49 was obtained between the scores on the addition facts test and the problem solving test. Correlations of .68 and .60 were obtained for the children in the lowest level and lowest IQ group respectively, between the same two tests. The correlations for the other levels and IQ groups were not as high, but some were significant.

Correlations of .46 and .41 were obtained between the scores of the problems with accompanying aids and the scores on the addition facts test and between the scores of the problems with no accompanying aids and the scores on the addition facts test. Correlations of .65 and .56 were obtained between the same three tests, respectively, for children in Level 4, and .52 and .54 for children in IQ group 78-100. The correlations for the other levels and IQ groups were lower, but some were significant.

The above correlations were all significant ($p < .01$).

Excellent prediction of the relative success in solving addition problems and learning the addition facts can be made for children entering the first grade. There are three categories of children for which it can be justified that the types of experience presently being provided produce different results with respect to solving addition problems. There is need for further research as to the type of arithmetic curriculum which would be most suitable for children in each of these three categories. Drill procedure on addition facts is quite ineffective for those children who experience difficulties in solving problems.

I
INTRODUCTION

Many present day arithmetic programs use Set Theory as a foundation on which to base the learning of natural numbers and operations on natural numbers.[1] Even though other mathematical approaches to the natural numbers exist, they have not been as amenable to the elementary arithmetic curriculum as Set Theory. Attempts are now being made, however, to construct arithmetic programs on other theoretical foundations,[2] but these materials are still in experimental form. Since the children who participated in this study were in a curriculum based on Set Theory, the terminology of sets only will be made clear in the next few paragraphs.

MATHEMATICAL BACKGROUND

A set, even though "set" is an undefined term,[3] may be thought of as "...formed by the grouping together of single objects into a whole."[4] Or, equivalently, a set is a "well defined collection of objects...."[5] The operation on sets which is of concern in this study is the operation of union. The union of two sets, A and B, denoted by $A \cup B$, "...is the set of all elements which belong to A or to B or to both."[6]

Two sets, A and B, are said to be equivalent if and only if they can be placed in one-to-one correspondence.[7] Two sets, A and B, are said to be in one-to-one correspondence when there exists a pairing of the elements of A with the elements of B such that each element of A corresponds to one and only one element of B, and each element of B corresponds to one and only one element of A.[8] The successor set of a given set A is the set which includes the members of A as well as A itself.[9]

Using the above concepts, the natural numbers may be constructed.[10] First the number 0 is identified with $\{\ \} = \emptyset$ that is, the set with no members. The number 1 is identified with $\{\emptyset\}$, that is, the set which contains the empty set. The number 2 is identified with $\{\emptyset, \{\emptyset\}\}$; the number 3 with $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$ etc. If the numbers are replaced for the respective sets, then $0 = \emptyset$, $1 = \{0\}$, $2 = \{0, 1\}$, $3 = \{0, 1, 2\}$, etc. The concept of the union of two sets is used implicitly in the above discussion since $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = \{\emptyset, \{\emptyset\}\} \cup \{\{\emptyset, \{\emptyset\}\}\}$, or $\{0, 1, 2\} = \{0, 1\} \cup \{2\}$, etc. Any set equivalent to $\{0\}$ has the natural number 1 associated with it; so in effect 1 is the standard name of a set of equivalent sets as are 2, 3, 4, ...

The sum of two natural numbers may now be defined by making use of the union of two disjoint sets; that is, two sets with no common members.[11] For this purpose let N(R) be the notation for the natural number identified with R. Then, for any two disjoint sets R and S, $N(R \cup S) = r + s$.[12]

There are many different substantive approaches to constructing an arithmetic curriculum based on the use of sets among which is the intuitive set approach.[13]

In the intuitive set approach, sometimes called the environmental approach,

> ...children learn first how to observe collections or sets of objects in the room, how to construct a set of objects, and how to describe a set. Next they match two sets of objects, by pairing elements, to discover which set has more, fewer or the same manyness of objects. It is essential at the start that the learners recognize the conservation of a set regardless of the arrangement of the elements. Their recognition is further strengthened by making numerous rearrangements of each set, and numerous *mappings* in which the more, fewer, or the same remain constant....
>
> In all this work there is no symbolism except the numerals, such as 3, 5, 2, ... which are the number names. In fact, such symbolism as $n\{\Delta, \square, \square\} = 3$ is a formalism, unnecessary and perhaps a hindrance to the intuitive grasp of number. For later work in problem solving, which involves the recognition of a concept in a real situation, and the building of a model of related mathematical concepts to interpret a

physical situation, it may be that the intuitive... approach to first learnings is indeed the best.[14]

To learn the operation of addition, the children study two separate sets of objects and then study a new set of objects formed by combining the two separate sets. The transition is then made to pictorial material where the children must participate at a somewhat higher level because the combining or separating of the two sets does not actually occur. For the abstraction, the numeral is presented along with the pictorial representation, and the pictorial is gradually eliminated.[15]

The intuitive set approach is exemplified for the first grade arithmetic program by the series Modern Arithmetic Through Discovery.[16] This is the arithmetic program used by the school system that participated in this study. Through physical objects and pictorial representations the idea of set—one-to-one correspondence, more than, as many as, fewer than, natural numbers and counting—are developed. After the natural numbers 1--9 have been developed and ordered, the sum of two natural numbers is developed. Again, this is done by means of physical objects and pictorial representations and is based on the concept of set union. However, other than merely developing sums, the program stresses number stories and number sentences.[17]

PSYCHOLOGICAL CONSIDERATIONS

In The Child's Conception of Number, Jean Piaget related the following experiment: "Bes, age 6 years, 2 months: (the experimenter had drawn a picture of 12 girls and 2 boys). Are there more girls or more children ?--More girls--Why ?--there are only two boys--But are girls children ?--Yes--then are there more girls or more children ?--More girls."[18] This experiment exemplifies two goals that Piaget has in this book: (1) to demonstrate stages in the development of particular concepts, and (2) to demonstrate the development of a conceptualizing ability that underlies the formation of any particular concept. For Piaget there are four main stages in the development of this conceptualizing ability: 1) Sensory-motor, preverbal stage; 2) Preoperational representation; 3) Concrete operations; 4) Formal operations.[19]

For the purposes of this study, the first and the fourth stages are not of much interest and consequently will not be discussed. The second stage, which occurs in children approximately between two and seven years of age, can be characterized by the beginning of thought, or preoperational representations. According to Piaget,

> ...an operation is an interiorized action... in addition, it is a reversible action; that is, it can take place in both directions....
>
> Above all, an operation is never isolated. It is always linked to other operations, and as a result, it is always a part of a total structure....
>
> To understand the development of knowledge we must start with an idea which seems central to me....the idea of an operation. Knowledge is not a copy of reality. To know an object, to know an event, is not simply to look at it and make a mental copy or image of it. To know an object is to act on it. To know is to modify, to transform the object, and to understand the process of this transformation, and as a consequence to understand the way the object is constructed. An operation is thus the essence of knowledge; it is an interiorized action which modifies the object of knowledge. For instance, an operation would consist of joining objects in a class. . . .In other words, it is a set of actions modifying the object, and enabling the knower to get at the structure of the transformation.[20]

If a child is at the stage of preoperational representation, Piaget continues,

> There is as yet no conservation, which is the psychological criterion of the presence of reversible operations. For example, if we pass liquid from one glass to another of a different shape, the pre-operational child will think there is more in one than in the other. In the absence of operational reversibility, there is no conservation of quantity.[21]

Moreover, Piaget further postulates that conservation of something is a necessary condition for any mathematical thought.

> Our contention is merely that conservation is a necessary condition for all rational activity. . . .This being so, arithmetical thought is no exception to the rule. A set or collection is only conceivable if it remains unchanged irrespective of the changes occurring in the relationships between the

elements. . . . In a word, whether it be a matter. . . of sets and number conceived by thought . . . or of the most refined axiomatization of any intuitive system, in each and every case conservation of something is postulated as a necessary condition for any mathematical understanding.[22]

This postulate has been verified in at least one empirical study conducted by Smedslund.[23] In this study he used 160 children in an age range of four years three months to eleven years four months. The children were fairly evenly distributed by age and sex. He found that the conservation of discrete quantity preceded transitivity of discrete quantity in all but four cases. He commented that in at least two of the four cases, the transitivity tasks involved a relative facilitation and hence could have produced the exceptions.

The third stage, that of concrete operations, is the stage in which the first operations occur. Piaget calls these concrete operations because ". . . they operate on objects, and not yet on verbally expressed hypotheses. For example, there are the operations of. . . elementary mathematics, of elementary geometry."[24]

Piaget's first goal in The Child's Conception of Number is to demonstrate stages in the development of particular mathematical concepts. These stages are: 1) Absence of conservation; 2) Intermediary reactions; 3) Necessary conservation.[25] Perhaps the best way to characterize these stages is to recount experiments reported by Piaget.

> (Stage 1) Co, five years of age. Like some of the other children at the first stage, Co took the counters one by one, thus apparently carrying out an operation of correspondence belonging to a higher level, but this was really not so, as we shall see. Having divided the eighteen counters, one at a time, into two heaps of nine, he was not certain that the two halves were equal! "Have I got the same amount as you?"-- (He looked at the two heaps, which were of slightly different density, then tried to make them the same.) -- But how did you divide them? . . . Are they the same? No-- How do you know? -- (He rearranged the heaps.). . . .
>
> (Stage 2) Pi, five years one month of age, with a total set of eighteen, took one counter after another and put them into two subsets, making a mistake of one unit, the result being ten and eight. He then rearranged each subset as a row of pairs, and compared the lengths of the two rows. Then, he spread out the pairs of the row of eight so that it was the same length as the other, but seeing the difference in density, he took one counter from the ten and added it to the eight, so that he had two similar sets of nine. "Are they the same?"--Yes-- (the elements of A_1 were then arranged as two rows, six and three.) Are they still the same? -- No -- Why? I've got more. ($=A_2$ the unchanged figure). . . .
>
> (Stage 3) Dre, six years and ten months of age, divided eighteen counters one or two at a time into two sets of nine, and was sure that they were equal even when the distribution was changed.[26]

On Stage 1, Co does not exhibit the property of conservation of numerousness. By this I mean, irrespective of how a set of objects is rearranged, the number of objects remains the same. Or, equivalently, if two sets are in one-to-one correspondence, then the number of objects in each is the same, regardless of the arrangement or rearrangement of the objects. On the second stage, Pi set up a correspondence, but it was destroyed by a rearrangement of one of the sets. On the third stage, Dre exhibited the property of conservation of numerousness. Piaget says of these stages: "The ordering. . . is constant, and has been found in all the societies studied. . . . However, although the order of succession is constant the chronological ages of these stages vary a great deal."[27]

Coxford gives an outline for the approximate age range for the attainment of number concepts. A first age range (Stage 1) is from four and one-half to five years. A second age range (Stage 2) is from five to six years. A third age range (Stage 3) is from six to seven and one-half years.[28] This by no means implies that if a child is from six to seven and one-half years of age that he has a good chance of being on Stage 3 in all of Piaget's tasks.

PROBLEM SOLVING

After reviewing the mathematical background a substantive approach and some psychological considerations of the attainment of number and related concepts, the next thing to be reviewed is problem solving in mathematics at the elementary school level. The meaning of "problem" as it will be used here is: "A quantitative situation, described in words, in which a quantitative question is raised without an accompanying statement as to the arithmetical operation required."[29]

Howard Fehr gives, as one among four basic outcomes of mathematical learning at the elementary school level, the acquisition of basic concepts and the application of the concepts to problem solving.[30] With reference to problem solving, he goes on to state, "There is no known system which the mind can practice for developing problem solving ability...; there are various hypotheses on the manner in which problems are solved but certainly without a host of well developed concepts, it is very unlikely that a problem can be solved. ..."[31] Henry Van Engen has this to say about problem solving.

> ...we want the child to grasp the structure of the problem before he looks for the answer. . . . The basic difference between good problem solvers and poor problem solvers must reside in differences in ability to recognize the element which we have called structure.... The method of problem solving we have illustrated here is a mathematician's approach to problems in miniature. One first searches for the fundamental structure of the problem situation; then he finds the appropriate symbols to express this structure....Certainly no "cue" method or mere admonition to think hold the mathematical power that the search for the structure of the physical situation can command.[32]

It seems then that two necessary conditions for solving an arithmetic problem are: 1) To have the arithmetical operation(s) required to solve the problem well mastered. 2) To be able to recognize what arithmetical operations are relevant to the problem (if any).

II

BACKGROUND OF THE PROBLEM

EMPIRICAL STUDIES

Even though statistical studies which are replications of the studies performed by Piaget have detected variability in stages of conservation of numerousness, these same statistical studies have also supported the existence of the stages.

In his first study replicating Piaget's experiments, Elkind gives the following summary:

> Eighty...children were divided into three age groups (4, 5, 6-7) and tested on the three Types of Material for three Types of Quantity in a systematic replication of Piaget's investigation of the development of quantitative thinking. Analysis of variance showed that success in comparing quantities varied significantly with Age, Type of Quantity, Type of Material and two of the interactions....
>
> The results were in close agreement with Piaget's finding that success in comparing quantity developed in three, age related, hierarchically ordered stages....[33]

The types of material Elkind used were 1) wooden sticks 1/4" square by 1 1/4", 2) orange colored water, a tall narrow glass, and two drinking glasses, one a 16 ounce glass and one an 8 ounce glass, and 3) large wooden beads that would just fit into the tall narrow glass in (2) above. The types of quantity he compared were 1) gross quantity, 2) intensive quantity, and 3) extensive quantity.

In explanation of these types of quantity, Piaget says,

> ...the question to be considered is whether the development of the notion of conservation of quantity is not one and the same as the development of the notion of quantity. The child does not first acquire the notion of quantity and then attribute constancy to it. ...At the level of the first stage, quantity is therefore no more than the asymmetrical relations between qualities, i.e., comparison of the type "more" or "less" contained in judgments such as "it's higher," "not so wide," etc. These relations depend on perception, and are not as yet relations in the true sense, since they cannot be co-ordinated one with another in additive or multiplicative operations.[34]

Elkind thereupon defines gross quantity as "single perceived relations between objects (longer than, larger than) which are not coordinated with each other."[35]

Intensive quantity is "the name given to any magnitude which is not susceptible of actual addition, as for example temperature. Two quantities of water at 15° and 25° respectively do not produce a mixture at 40°."[36] Also, the relations Piaget talks about when describing gross quantity begin to be coordinated at the second stage and "result in the notion of intensive quantity, i.e., without units, but susceptible of logical coherence."[37]

Extensive quantity is "the name given to any magnitude that is susceptible of actual addition, as for example mass or capacity...."[38] "As soon as... intensive quantification exists, the child can grasp...the proportionality of differences, and therefore the notion of extensive quantity."[39]

In the study, gross quantities were easiest to compare, intensive were intermediate, and extensive were hardest. For the types of material, quantities involving liquids were hardest to compare, with no difference between sticks and beads. There was a significant interaction of age groups and the quantity compared. Comparisons involving gross quantities was easy for all three groups. However, comparison involving intensive quantities was quite difficult for the 4-year group and became increasingly easier for the two older groups. The same was true for comparisons involving extensive quantities, but these comparisons remained more difficult than the comparisons involving intensive quantities.

Since Piaget defines his stages in terms of the type of quantitative comparisons children are capable of making,[40] it is clear from Elkind's study that a child may be able to make extensive quantity comparisons using materials of a given kind and thereby be classified at Stage 3, but changing the type of material could affect the type of quantity comparison the child is capable of and thereby alter the stage classification. However, there is a definite statistical relationship between age groups and stages as exemplified by the interaction of age groups and quantity compared and high and significant correlations between types of material.

Dodwell also has observed variability of stages. He studied—using 250 children in an age range of about 5 to 8 years—1) relation of perceived size to number using beakers and beads, 2) provoked correspondence using eggs and eggcups, 3) unprovoked correspondence using red and blue poker chips.[41]

In the first category above, about twenty-five per cent of the children at 6 years 2 months of age showed Stage 3 responses, in the second category, about sixty per cent of the children at 6 years 2 months showed Stage 3 responses, and in the third category, about twenty per cent of the children showed Stage 3 responses. In this same study he observed a low but significant correlation (-.24) between IQ and Stage 3 responses indicating that intelligence is a factor in conservation problems. The name of the IQ test was not given.

Van Engen and Steffe[42] in a study involving 100 first grade children have also observed a low (.24) but significant correlation between IQ (as measured by the Kuhlmann-Anderson Intelligence Test) and the success of the children in four tasks involving concepts in addition.

Dodwell and Elkind have also performed replications of Piaget's experiments on the ability of children to include partial classes within a total class, i.e., if $A \cup B = C$ $(A \cap B = \emptyset)$, then $A \subset C$ or $B \subset C$. For his subjects, Elkind selected twenty-five children from each of the grades kindergarten to third.[43] The question asked of each child was, "Are there more boys (or girls depending upon the sex of the child being questioned) or more children in your class?"[44] Other questions were also asked to gain assurance the children understood the above question. On the basis of the responses, the children were placed in three stages; Stage 1 if either $C \subset A$ or $C \subset B$, (A = boys, B = girls, and C = children), Stage 2 if $C = A$ or $C = B$, and Stage 3 if either $C \supset A$ or $C \supset B$. A χ^2 performed on age groups by stages was significant. Fifty per cent of the five-year olds, thirty-two per cent of the six-year olds, twelve per cent of the seven-year olds, and eight per cent of the eight-year olds were in Stage 1. Correspondingly 48, 56, 76, and 92 per cent respectively were in Stage 3.

Dodwell was interested in investigating the responses to class inclusion questions and responses made on the tests of provoked and unprovoked correspondence discussed earlier.[45] In the discussion of the results, he states that the "ability to answer correctly questions which involve simultaneous consideration of the whole class and its (two) component subclasses, appears to develop to a large extent independently of understanding of the concept of cardinal numbers (as measured by the tests for provoked and unprovoked correspondence...."[46]

The above studies are what may be called "one-shot" studies, that is, studies that test an individual at a point or points in time. The question immediately arises then, if a child is on a given stage at a given point in time with reference to a particular situation and particular materials, will the same child be on the same stage at a different point in time, all other things constant? Dodwell, using the tests devised in an earlier study,[47] made a test-retest reliability study with intervals of one week and three months. He comments, "The short term reliability of the test is highly satisfactory, and compares well with the reliabilities of commercially available cognitive tests. The long term reliability indicates considerable stability in the development of number concepts...."[48]

In this same study, Dodwell examined the data from his original sample of 250 children to detect differences due to sex and socio-economic status. He reports that "Differences were extremely small, insignificant, and did not favor either sex."[49] To test for socio-economic status, the children were divided into three groups on the basis of their fathers' occupations: 1) professional, 2) clerical and semi-skilled, and 3) semi-skilled or unskilled trades. No differences were detected among the groups, but the higher socio-economic groups scored more favorably.

Van Engen and Steffe also have observed no differences between first grade boys and girls when studying addition concepts.[50] These children were taken from five schools all ranked as serving a middle-class population by school officials. There was one school in which the children did consistently better than in the four other schools, but not significantly better.[51]

The difficulty of categorizing population of children into three distinct stages notwithstanding, the relationship between success of children in arithmetic and the three stages of Piaget has not been made explicit. However, Dodwell has worked on this problem.[52] He gave forty kindergarten children his number test in the final term of the school year. He then followed thirty-four of these children through the first term of the first grade in that a group of first grade teachers constructed a test involving number recognition, counting, drawing different numbers of objects, etc. which "seemed to the writer to be a fair test of the curriculum covered."[53] A correlation of .59 was obtained between this test and the test scores obtained previously. The interval between the two was about seven months.

STATES AND TRANSFORMATIONS

It has been noted that Piaget says that mental operations are interiorized actions. With regard to these interiorized actions, Piaget states, "Two quite different aspects of knowing come into play, depending on whether we are dealing with states or with transformations leading from one state to another."[54] He goes on to say that physical actions which transform objects in any way and interiorized actions are examples of transformations. The connections between the physical actions and the interiorized actions are made clear by Piaget.

> First of all, there is what I call physical experience and what I call logical-mathematical experience. Physical experience consists of acting upon objects and drawing some knowledge about the objects by abstraction from the objects. (Weight, for example)...; there is the second type of experience which I shall call logical-mathematical experience where the knowledge is not drawn from the objects, but is drawn from the action effected on the objects. This is not the same thing. When one acts upon objects, the objects are indeed there, but there is also the set of actions which modify the objects.[55]

So, physical action and interiorized actions are both examples of transformations on a set of objects. However, once a set of objects is transformed physically, at least three possibilities then arise. The first is that the child looks upon the transformation as two states, the original and the final state of the objects, and ignores the transformation. He is thus unable to draw any knowledge from the transformation. His experience with the objects then, is of a qualitative type and even though he says there are, for example, five objects in the set in its original state and five objects in the set in its final state, the fiveness associated with each state is not connected, and when asked to compare the numerosity of the objects in the original and final states, he will be unable to do so correctly.

The second possibility is that the child does not ignore the physical transformation effected on the objects and knows a priori that the "fiveness" associated with each state is the same "fiveness." The action in this case is not external to the child but has been interiorized and has meaning.

A third possibility that arises is that the child does not totally ignore the transformation but neither does he know a priori that the "fiveness" is the same in both states. That is, he makes judgments that are inconsistent with each other. As noted, Elkind and Dodwell have both pointed out that both the materials and situation may influence the judgments of children in conservation tasks.

The application of the above discussion to first grade children's learning addition has been studied by Van Engen and Steffe. They state, "The ability on the part of the child, to respond correctly to an addition combination, for example 2 + 3, seems to have little or no relation to his ability to ignore his perception when two groups of objects, one of 2 and one of 3, are physically transformed into a group of 5."[56] The most feasible interpretation of this phenomena is that, in their words, "the children have not abstracted the concept of the sum of two whole numbers from physical situations but have memorized the addition combinations."[57]

The children, then, may look upon the two groups as one state, the combined group as another state, and not be aware of the connection (that is, the physical transformation) between the two states and thereby be unable to make correct comparisons between the two states when asked to do so.

The idea of a physical transformation is used in elementary arithmetic programs to teach children addition, subtraction, multiplication, and division. For example, in the series, Seeing Through Arithmetic the operation of addition is presented to first grade children by means of joining one set of objects with another. The authors describe the situation in this manner.

In the illustration below, 4 blocks are being pushed toward 3 blocks, implying a joining action. This situation is additive and is symbolized mathematically as $3 + 4$. . . . The "3" gives the number of blocks in the original set, the "4" tells how many blocks are in the joining set, and the "+" is used when a joining action occurs. The phrase $3 + 4$ is a name for the total number of objects in the set. The phrase $3 + 4$ names the number and also tells what is happening in the physical situation. . . .

Pictures that show additive action. . .help to develop an intuitive understanding of what addition and subtraction mean. However, the operations of addition and subtraction and the phrases that correspond to these actions deal with numbers, not objects. . . .

Children think in terms of action—that is, in terms of what they have seen happen or what they have caused to happen to objects. In the. . . program, a wealth of opportunities is provided for the children to work with situations in which they see the joining of sets of objects. . . . Pictures are used to show action situations. It is important for the children to learn to "read" and interpret their pictures because at this level the pictures are used in place of printed words.[58]

III
THE PROBLEM

THE BASIC PROBLEM

As noted, Dodwell has made a correlation study of number comprehension tests based on Piagetian tasks and success on a test purportedly constructed to measure achievement in arithmetic. The substantial correlation of .59 obtained gives encouragement for a more carefully designed study to be undertaken to ascertain the performance of children who have been categorized into groups by a test based on Piagetian tasks, on a particularly important aspect of first grade arithmetic.

It is well known that the ability to solve problems is one of the most important outcomes of the arithmetic curriculum. Because of the importance placed on problem-solving and because addition is introduced by means of problem-solving, this study is primarily concerned with the performances of first grade children (categorized into levels by a pretest administered immediately before a test of the ability to solve problems) in a uniform arithmetic curriculum, when solving problems with an additive structure.

ARITHMETIC PROBLEMS INVOLVING A TRANSFORMATION

An arithmetic problem has an additive structure if it is an instance of the union of two or more sets. For example, the problem "John has three apples and Mary has 4 apples. How many do both children have ?" has an additive structure because the problem involves a set of 3 apples, a set of 4 apples, and the union of the two sets or the set of 7 apples. The problem "There are 4 dogs on a rug. Seven more join them. Now how many dogs are on the rug ?" also has an additive structure but differs from the first problem in an important dimension. In the first problem, both sets are static. Neither set has the possibility of movement and the union of the two sets is implied by the question only. In the second problem, the movement of the set of seven dogs is described in the problem. However, both problems do involve set inclusion.

Piaget, while studying set inclusion, concluded,

> All the children quoted understood the nature of the sets involved in the problems of inclusion. . . .These children were. . . clearly conscious of the general definition of the total set in question. . . .It cannot therefore be disputed that all these children possessed the notion of the total class required by the questions and were capable of the general statement defining that class. . . .
>
> And yet, as soon as it becomes necessary to think simultaneously of the whole and the part, as our question requires, difficulties arise. The child apparently forgets the whole when he thinks of the part, and forgets the part when he thinks of the whole. Or rather, when he thinks of the whole, he can envisage the parts which have not yet been dissociated, but when he tries to dissociate one of the parts he forgets the whole. . . .In other words, the children quoted above cannot establish a permanent inclusion between the whole and the parts: as soon as the whole is divided, even in thought, the parts cease to be included in it and are merely juxtaposed without synthesis.[59]

As previously noted, Elkind has verified the above results, using kindergarten to third grade children.

In the first problem given above, the children must think simultaneously of the whole, the apples they both have, and also of the two parts, the apples each has. This is also true of the second problem. The children must think of the whole, the dogs on the rug, and the parts, the dogs originally there and those that joined. However, while both problems involve two states, the second problem involves a described transformation from one state to another. It may be true then, that while the children are not able to supply their own transformation from one state to another, and have difficulty thinking about class inclusion, the described transformation may connect the states for the children and hence alleviate the problem of class inclusion.

THE PRESENCE OR ABSENCE OF AIDS WHEN SOLVING ARITHMETIC ADDITION PROBLEMS

In the intuitive set approach to learning addition, it has been noted that children first study addition by means of sets of physical objects, then proceed to pictorial representations, and finally get to more abstract situations. Moreover, the children are encouraged to verbally relate the number stories that go with the problem in order that the problem-solving abilities of the children will become independent of visual aids. There are empirical studies which do not support this approach.

Piaget calls the first operations concrete because "...they operate on objects, and not yet on verbally expressed hypotheses."[60] In a study on concrete reasoning, Jan Smedslund has this to say as some of his concluding remarks.

> ...the data are not inconsistent with the hypotheses that perception tends to be subordinated to and aid reasoning more frequently in subjects above than in subjects below seven years six months and that perception tends to disturb or at least not aid reasoning more frequently in subjects below seven years six months than in subjects above that age. Although the data are restricted and not entirely compelling, they invited certain speculation about the role of perceptual materials, for example in the teaching of elementary arithmetic, which definitely involves the acquisition of concrete reasoning. The findings presented here certainly do not support the hypotheses that logical reasoning first occurs with perceptual support and then gradually begins to function in the absence of such support. On the contrary, it is possible that early training should be given mainly in situations without perceptual support.[61]

These conclusions are in direct conflict with the intuitive set approach to learning addition at the first grade level.

In a study of the performance of second grade children on four kinds of division problems, Marilyn Zweng observed that "scores for problems requiring a drawing (the subject was required to make a drawing before solving the problem) were in general lower than the other scores (problems given with accompanying physical objects for the subject to use to solve the problem)."[62]

While the conclusions of these two studies do not conflict due to the age differences of the children, based on Smedslund's conclusions Zweng may have observed different results at the first grade level. Moreover, in Zweng's study no attempt was made to categorize the children on any level of conservation of numerousness, so that no conclusion could be drawn for second grade children relative to Smedslund's comments.

TESTS OF CONSERVATION OF NUMEROUSNESS

Relative to the basic problem of the study, a test of conservation of numerousness was developed based on the types of quantitative comparisons that children are known to make, that is, gross quantitative comparisons, intensive quantitative comparisons, and extensive quantitative comparisons. The tests Dodwell constructed, which have already been discussed, were direct replicas of those used by Piaget to assess number concepts in young children. There are five different situations on which the children were tested and on which they performed quite differently. Dodwell reports that the probability of a child's making extensive quantitative comparisons in the case of unprovoked correspondence, given that he has correctly answered questions relative to cardinal and ordinal numbers, was .8. Moreover, if a child did not respond correctly to questions involving cardinal and ordinal numbers (was not operational), then the probability of his making extensive quantitative comparisons in the case of unprovoked correspondence was .06. Because of this high probability (.8), situations that can be classified as unprovoked correspondences were selected for use in this study. Dodwell's test could not be directly utilized since it involved subtests which were not directly concerned with the conservation of numerousness. Moreover, the test on unprovoked correspondence involved a physical transformation. When Dodwell reported the results of a group paper and pencil test which he constructed based on his original test, he concluded that,

> Although the group test measures under-

standing of number and related concepts in situations apparently similar to those used in the individual test, it is arguable that it in fact measures a different aspect of the child's cognitive abilities. In the individual test one is measuring ability or understanding when the child actually perceives the transformation on the test materials; in the group test, on the other hand, the child is faced with fixed alternatives between which he has to choose, and therefore has to imagine the transformation....[63]

Since a long-term objective is to construct a group paper and pencil test of conservation of numerousness, it is apparent that quantitative comparisons that do not involve a physical transformation are most amenable to that objective.

Other tests have also been developed to test conservation of numerousness, among which are those developed by Almy,[64] Churchill,[65] Elkind,[66] Feigenbaum,[67] Smedslund,[68] and Wohlwill,[69] all of which involve a physical transformation of the elements.

Lunzer reports that,

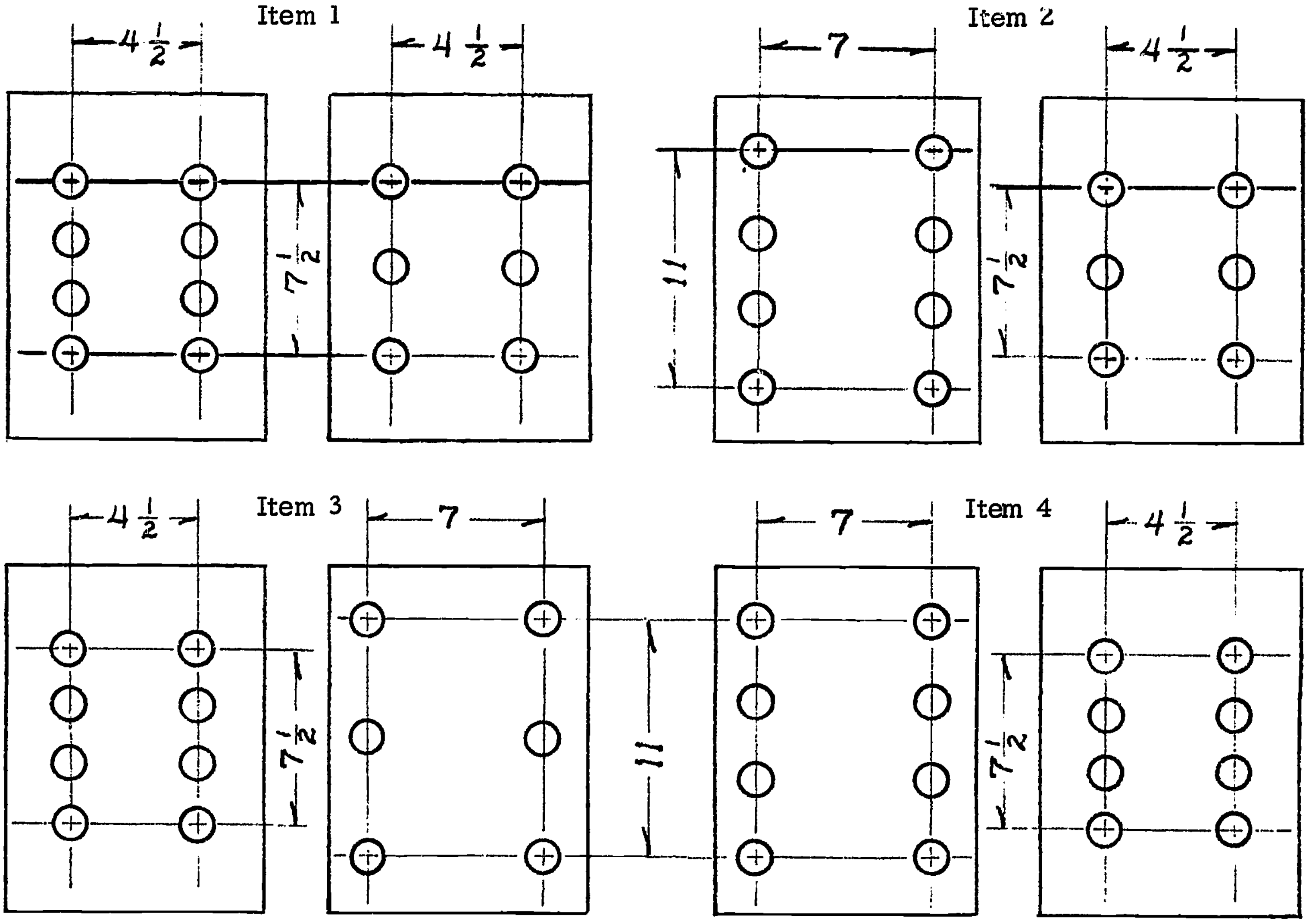
> Phemister has recently shown that when children are shown two circular arrangements of beads of which one is larger so that the beads are more spaced, but contain only eight beads while the other, tighter bracelet has nine, only a minority of children before the third year of the junior school [third year of junior school in England is equivalent to the fourth grade] are willing and able to count the beads in each set. . . . Children who failed in applying their understanding of conservation and number when the situation was difficult had no difficulty recognizing these concepts in... easier situations of Piaget.[70]

Following Lunzer's comment, it seemed quite reasonable to start with circular arrangements of objects when constructing the test of conservation of numerousness using situations that are classifiable as "unprovoked correspondence." Since Lunzer indicates that the circular arrangement was quite difficult for children before the second grade, it was decided to use two more geometrical arrangements. Figures 1, 2, and 3 depict the tests. The question for each item of each test was the same as follows: "Are there more______ here (the experimenter pointed to one of the collections) or are there more______here (the experimenter pointed to the other item) or are there the same number of______here as here ?" The word "more" was selected rather than the word "fewer" because in their analysis of reasons for incorrect responses, Van Engen and Steffe[71] observed that no child gave the response "fewer" when asked why they selected candies either in two piles or in a combined pile. The use of the word "more" was very frequent. Moreover, Elkind[72] used the terminology "the same number" in one of his replication studies and assumed that such terminology did not provide clues to the type of quantity that had to be compared. Smedslund, too, when studying conservation of numerousness has used a question similar to that used in the present study.[73] This terminology is also used in the intuitive set approach to learning arithmetic.

Through the intuitive set approach, it was possible for the children to have at their disposal the transformations of 1) one-to-one correspondence and 2) comparison by counting. The other types of comparison that were possible are: 3) guessing (no comparison), 4) comparison by relative sizes, and 5) comparison by relative density of the objects. The first and second methods of comparison are comparisons of extensive quantity. The fourth and fifth methods of comparison are comparisons of gross quantity. A combination of 4 and 5 can be thought of as a comparison of intensive quantity. An example will illustrate the types of comparisons.

In Test 3, Item 1, the children are asked to compare two circles of blocks, each four inches in diameter. One has six blocks on its perimeter and the other has eight. Here, if a child judges on relative size alone (making a gross comparison) he will no doubt get the item wrong. Also gross comparisons can be made on relative density alone. One circle is more dense than the other, so it has "more," a "correct" response. A child may also make an intensive judgment that the one circle has more blocks because both circles are both the same size and one is more dense, also a correct judgment. If a child sets up a correspondence between the two sets or counts each set, he has made an extensive comparison.

Therefore, it is possible for a child to respond correctly on this item without making an extensive quantitative comparison. The same can be said for Items 2 and 3. In Item 2, the circle with 8 has a four-inch diameter and the circle with 6 has a seven-inch diameter. A child could make a gross comparison based on density and respond correctly on the item, or he could make an intensive comparison



Note: All dimensions in inches.

Fig. 1

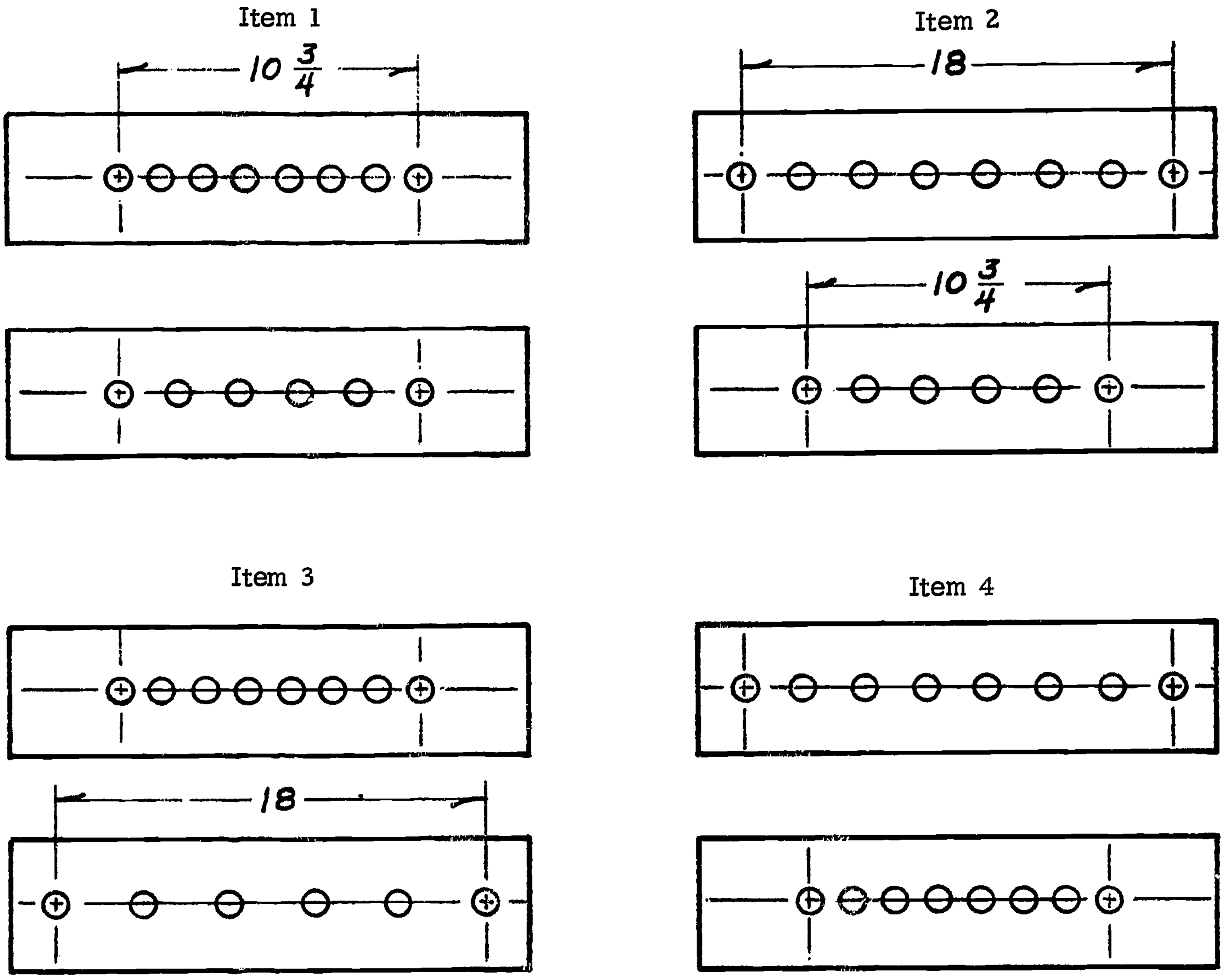
Test 1. White styrofoam balls about 1 1/2" in diameter arranged on orange construction paper.

based on the fact that the circle with a four-inch diameter is smaller but the objects are closer together than are the six objects relative to the circumference of the circles. A similar discussion can be given for Item 3. However, in order for a child to respond correctly on Item 4, an extensive comparison must be made, disregarding guessing. Since the two circles are of the same number but of different diameters, any gross comparison will lead to an incorrect result. An intensive comparison is impossible since there are the same number of blocks in each circle, all equally spaced, so that the distance between the blocks is always in the same proportion to the diameter. A similar discussion can be given for each item of Tests 1 and 2.

The three tests are isomorphic in structure but differ in the materials used and the geometric configuration. Three tests were constructed because it was felt, considering Lunzer's comments, that a much more reliable estimate could be made of the children's ability to make extensive comparisons than if only one test were used. Moreover, the geometrical configuration and materials were varied because of Elkind's and Dodwell's discovery that the type of materials used and the particular situation involved may affect the quantitative comparisons made.

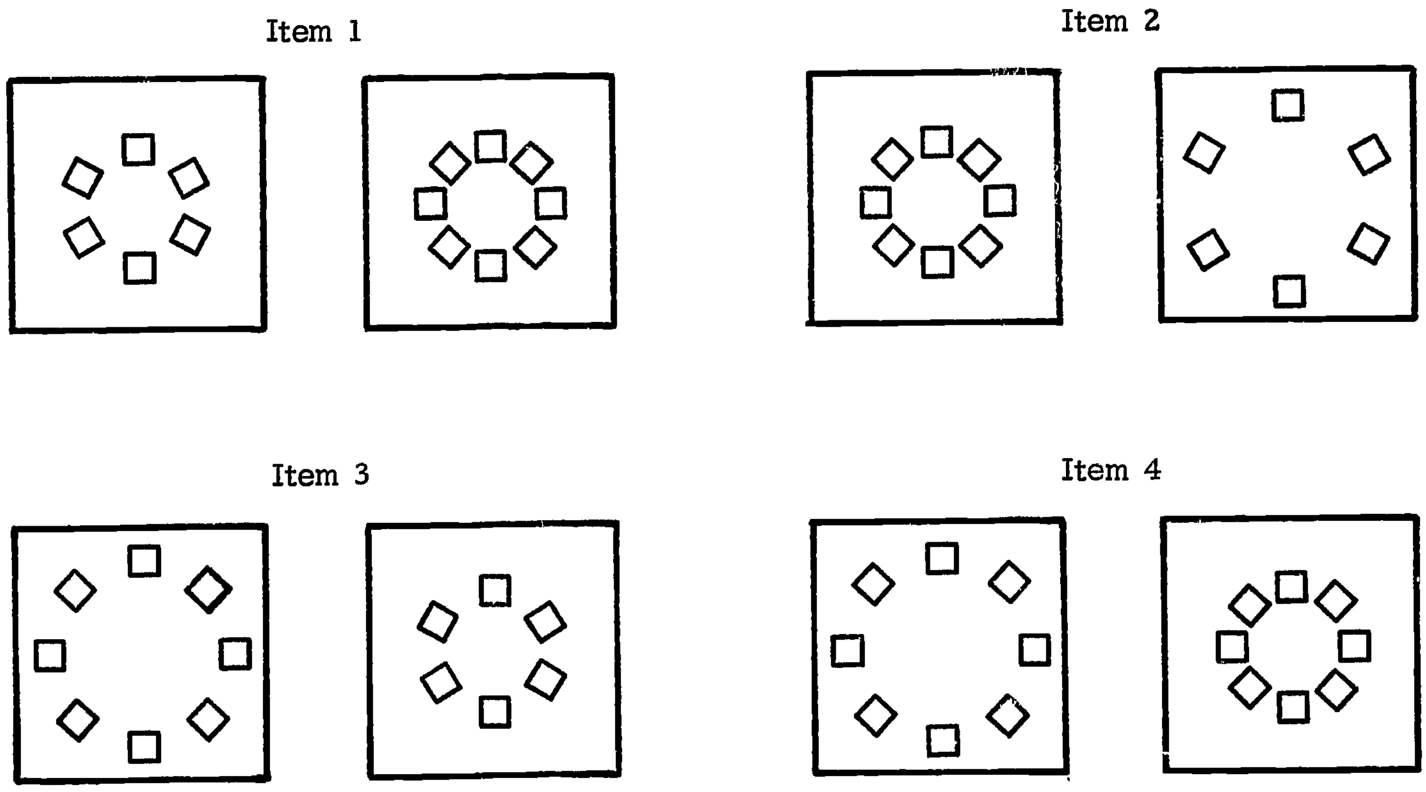
In Test 1, Items 1, 2, and 3, the eight styrofoam balls were held constantly on the left with size varying. In Test 2, Items 1, 2, and 3, the eight checkers were always on the top with size varying. However, in Test 3, the number of blocks was interchanged in Items 1 and 2 but not 2 and 3, and the sizes of the circles were varied. This was done to eliminate any possible right-left response bias which may have occurred from Test 1.

Sets of 6 and 8 objects were used because experience has shown that many children of this age are able to intuitively recognize sets of four and sometimes even sets of five but cannot readily recognize sets of 6 and larger. The configurations used were more amenable to an even number of objects than to an odd number since an odd number of objects would upset the symmetry. The linear arrangement was chosen because it may be more conducive to a one-to-one correspondence for children. The rectangular arrangement was chosen as a variant of the two dimensional circular configurations.



Note: All dimensions in inches.

Fig. 2
Test 2. Black checkers of 1 1/4" diameter arranged on orange construction paper.



Note: Circular patterns have 4" and 7" diameters.

Fig. 3
Test 3. Black 1" wooden cubes arranged on orange construction paper.

THE PILOT STUDY OF THE TEST ON CONSERVATION OF NUMEROUSNESS

After the three tests were constructed, an operational definition of the levels of conservation of numerousness still remained to be made. A pilot study was conducted in early November, 1965, using 81 first grade children from a school using the arithmetic television program, "Patterns in Arithmetic," developed by the Research and Development Center, University of Wisconsin. "Patterns" uses essentially the same substantive approach as that described earlier so the children had many experiences setting up correspondence, etc.

Procedure

The tests were arranged on tables with boards placed vertically to prevent the subjects from observing more than one item at a time. Tests were assigned in random order to the children, but the items were taken in order. The child and experimenter walked from item to item. The items were not assigned at random because experience showed it was very difficult for the experimenter and child to move together in a random order around the table without causing a lot of distraction for the child. It was felt any gain made by randomization would be more than an offset by the confusion caused by a random order.

Discussion of Results

Table 1 shows that the children made more correct responses on Test 1 than on either Test 2 or 3. Table 2, shows that more children scored (1, 1, 1, 1) on Test 1 than on either Test 2 or 3 and more children scored (1, 1, 1, 1) on Test 3 than 2. These two observations are consistent with the earlier discussion that different materials and situations do, in fact, influence the type of quantitative comparisons that children are able to make. It can be hypothesized that the reason the children were able to do much better on Test 1 than on 2 or 3 was that the rectangular arrangement lends itself more readily to setting up a one-to-one correspondence because the children could conceivably compare, for example in Item 4, four rows with four rows and two columns with two columns to make an immediate judgment of equinumerousness. Item 1 of Test 2 warrants discussion. A reason for the relative difficulty of this item probably was that the children were more inclined to make a gross comparison of length instead of density, and in Item 1 of Tests 1 and 3, they made either a gross comparison based on density or

Table 1

Total Correct by Tests and Items

	Item			
Test	1	2	3	4
1	70	71	77	54
2	49	61	75	36
3	69	68	69	37

Table 2

Frequency of Response Patterns by Tests

	(1, 1, 1, 1)	(0, 1, 1, 0)	(0, 1, 1, 0)	(0, 0, 1, 0)	(0, 0, 1, 1)
Test 1	47	17	5	2	2
Test 2	21	22	9	8	5
Test 3	27	24	5	3	2
	(1, 1, 0, 1)	(1, 1, 0, 0)	(1, 0, 1, 0)	(0, 1, 1, 1)	(1, 0, 1, 1)
Test 1	1	1	1	1	3
Test 2	0	1	3	5	2
Test 3	6	5	6	1	1
	(0, 1, 0, 1)	(1, 0, 0, 0)	(0, 1, 0, 0)	(0, 0, 0, 1)	(0, 0, 0, 0)
Test 1	0	1	0	1	0
Test 2	2	0	1	1	1
Test 3	0	0	0	0	1

0 = Incorrect Item 1 = Correct Item
(Item 1, Item 2, Item 3, Item 4)

an intensive comparison.

Even though children are capable of making extensive quantitative comparisons (5 children scored (0, 1, 1, 1) for Test 2), they still may not be able to ignore their perception in some cases and are led to make gross quantitative comparisons. However, there are possibly 8 children who responded correctly on the last item and incorrectly on at least one other item on all three tests. Inspection of Table 2 shows that 8 children responded correctly to Item 4 and incorrectly on some other item for Test 1, 15 for Test 2, and 10 for Test 3. The greater frequency for Test 2 is reflected in that only 49 of the 81 children correctly responded to the first item while 70 and 69 respectively responded correctly to the same item in Tests 1 and 3.

Table 1 shows a possible learning effect for the sequence of items in Test 2. However, in view of the correct responses for the two other tests, and an investigation of the actual test, a much more plausible explanation for the increasing number of correct responses can be given. Item 1 has been already discussed. In Item 2, there was not a great deal of difference in the density of checkers in the two rows. Therefore, those children making a gross comparison based on density could very easily make a wrong judgment. However, for Item 3, the row of 8 was much more dense than the row of 6, so that a gross comparison based on density would give a correct answer. Table 2 supports this analysis as there are 8 children who scored (0, 0, 1, 0) and 3 who scored (1, 0, 1, 0).

Four levels of conservation of numerousness were defined based on the data of the above three tables, in particular Table 3. Level 1 was made up of children who responded correctly on each item of each test, Level 2 of all children who responded correctly on all the items of exactly two tests, Level 3 of all children who responded correctly for all the items on only one test, and Level 4 of all children who responded incorrectly for at least one item on each test. On the basis of chance responses, the probability of a child being at Level 1 is .001; at Level 2, .026; at Level 3, .241; and at Level 4, .732. So any child who guesses is almost certain to be in Level 3 or 4, with Level 4 having a much higher probability.

A child's responses at Level 1 were not necessarily based on extensive quantitative comparisons, but it is quite unlikely that he would make no extensive quantitative comparisons: Item 4 of each test permits neither an

Table 3

Frequency of Tests Entirely Correct

	Tests Correct							
	1, 2, 3	1, 2	1, 3	2, 3	1	2	3	none
Frequency	8	7	9	4	23	2	6	22

accurate gross comparison nor an accurate intensive comparison; the probability of a correct guess on Item 4 of all three tests is only 1/27. Also, the probability of getting the other 15 items correct on a basis of gross or intensive comparison or guessing would lower the 1/27 and thus decrease the child's chances of being at Level 1. It is impossible to give probabilities for a child making a correct response by gross comparison on the first three items of each test since concentration on length or density is the choice of the individual and may even change from item to item.

The largest probability that a child at Level 2 has of not making any extensive quantitative comparisons is 1/9 and, as in the above discussion for Level 1, the probability of getting the other items correct by gross or intensive comparison or guessing will lower this probability, but it is impossible to tell by how much. The largest probability a child at Level 3 has of not making any extensive comparison is 1/3, and again this figure should be lowered by consideration of the gross or intensive comparison or guessing made on the other items.

Level 4 children could have made extensive quantitative comparisons, but it is much more likely that the responses were based on either guessing or gross or intensive comparisons. Of the 22 children at Level 4, 14 of them had Item 4 wrong in all three tests. Of these 14, 11 had other items wrong. There is a high probability that these 14 children were responding on a basis of intensive or gross comparisons. Three of the 22 children had one of the last items correct, but missed a total of 9 of the 27 first three items of each test. This again indicates a high probability of responses based on gross comparisons or guessing, and it is likely that the children guessed the correct answer for Item 4 of some test. Five of the 22 children had 2 of the 3 last items correct. However, 26 of their 45 total responses on the first three items were incorrect, again indicating a high probability of guessing on the last items. One of the 22 children had all three of the last items correct but had 4 of the 9 responses on the first three items correct.

STATEMENT OF THE PROBLEM

In the case of second grade children, as noted, there are observations that division problems can be solved more readily when the children are given physical aids to which they may refer during solution of the problem than when the children are required to make a drawing. For younger children, there are observations which do not support the hypothesis that logical reasoning first occurs with perceptual support, then gradually begins to function in the absence of such support. In fact, the conclusion was drawn that early training should be given in the absence of such support. It certainly may be inferred that these contradictory conclusions may be a function of the age level of the children, since Piaget and others have confirmed that older children are able to make extensive quantitative comparisons more successfully than younger children, and the age of 7 to 8 seems to be a usual dividing point. This inference leads one to hypothesize that first-grade children will perform differently on problems with accompanying physical aids depending on the degree to which they are able to make extensive quantitative comparisons. Since perceptual support is not limited to physical objects, the same hypothesis may be formulated using accompanying pictorial materials. The hypothesis also may be formulated with reference to problems without accompanying perceptual support, since Smedslund indicated a need for this investigation.

Relative to these three levels of visual support—physical, pictorial and no support—different kinds of transformations are possible.

In the case of physical objects, a transformation is possible. In the case of pictorial objects, an indicated transformation is all that is possible. In both cases, however, the transformation may be described since the problem is verbally presented to the child. A description is the only possible transformation in the case of a problem given verbally to a child without visual support. It may then be hypothesized that children will perform differently on problems involving a transformation and problems that do not involve a transformation, depending on the degree to which they are able to make extensive quantitative comparisons. There are, then, two variables in solving arithmetic problems with an additive structure that are of interest: 1) the presence or absence of aids, and 2) the presence of a transformation in the problem (described, indicated, or actually present) or the absence of a transformation. Three levels of the first variable were identified: 1) the presence of physical aids, 2) the presence of pictorial aids, and 3) the absence of aids. Therefore, there are six problem types of interest: 1) problems with physical aids present and a physical transformation involved, 2) problems with physical aids present and no transformation involved, 3) problems with pictorial aids present and an implied transformation involved, 4) problems with pictorial aids present and no transformation involved, 5) verbal problems with a described transformation, and 6) verbal problems with no described transformation. It is to be noted that regardless of the type of transformation (physical or implied) there will always be the described transformation implicit in the verbal statement of the problem. A review of the literature shows no indication that these six problem types with an additive structure have been systematically studied, nor has there been any study of the relationship of the six types of problems to first-grade children's ability to make extensive quantitative comparisons.

Low but statistically significant correlations have been observed between the types of quantitative comparisons children are able to successfully make and their scores on group IQ tests. This indicates that IQ is a factor in conservation problems. But it also indicates that IQ tests measure factors other than the ability to make extensive quantitative comparisons. Moreover, a review of the literature indicates no determination of the relationship of IQ to 1) the presence or absence of a transformation, 2) the presence or absence of visual aids, and 3) the six types of problems.

For purposes of this study it was decided to partition first grade children on the basis of the four levels of conservation of numerousness and three IQ groups: 78-100, 101-113, 114-140. The lower bound on IQ was selected to eliminate the "educable mentally retarded"[74] children. With regard to these children, Klausmeier states, "To expect the majority of them to achieve well in an algebra class, to read at ninth-grade level, or to understand abstract ideas is unrealistic."[75] The upper bound on IQ was selected to eliminate the highly gifted children. Klausmeier states, "The child who is already consistently superior in most areas of school instruction or who promises to be is called a gifted student.... IQ is usually one criterion of giftedness."[76] Terman has given a general classification,[77] in which he interpreted an IQ of over 140 as near genius or genius.

The IQ of 140 represented the 98.67 percentile of the first grade population on the Kuhlmann-Anderson IQ test and an IQ of 78 represented approximately the 2.4 percentile of the population. Only 10 children were eliminated from the original sample, 4 below an IQ of 78 and 6 above an IQ of 140.

There are, then, twelve groups of first-grade children who are identified for the purposes of this study. Schematically, if S_i represents level i, $i = 1, 2, 3, 4$ and G_j represents IQ group j, $j = 1, 2, 3$, then (S_i, G_j) represents the set of the 12 possible groups. All of the children in each group used the same arithmetic workbook.[78] School officials judged at least the first five-eighths would have been covered by all the teachers in the system.[79] Sums which add to 1 through 8 have been developed in this time through the use of concrete objects (which the authors suggest to the teachers) and the pictorial material of the workbook. In the first five-eighths of the book there are 71 problems which have accompanying pictorial representations with an implied transformation, 25 which involve no implied representations, and 44 which require the children to draw pictures of objects to complete number sentences in an additive situation. There are a total, then, of 96 problems which involve a transformation and 25 which do not involve a transformation. Moreover, numerous opportunities are given for drill on combinations. In this same time, however, addends with two digits are being learned for which there are 27 exercises which can be classified as accompanying pictorial material with no transformation present. While the addition facts are being presented, the authors direct the teachers to tell number stories about the pictures and also have the children tell number stories and write number sentences for the number stories. This practice is started from the first lesson on addition in which the teacher is directed to tell number stories which the children can act out. This progresses until the children are asked to make up their own stories for the pictures and write the accompanying number sentence. When learning to add using two digit numerals, the children are still asked to interpret the pictorial situation. All the above instruction has been given using the intuitive set approach described earlier in this paper.

The questions to be asked after children have been through the above described curriculum in problems involving addition are these:

1) Is the mean performance of children different for the six described arithmetic problems involving an additive structure?

2) Is the mean performance of the children in each of the 12 groups different when solving arithmetic problems with an additive structure?

3) Are the group profiles of the 12 groups different relative to the means of the six tests of addition problems?

4) Is the mean performance of the children different in the four levels of conservation of numerousness?

5) Is the mean performance of children different for the problems involving the three levels of visual aids: Physical objects, pictorial objects, and no visual aids?

6) Is the mean performance of children different for the problems describing a transformation and the problems that do not describe a transformation?

7) Is the mean performance of the children different in the three IQ groups?

8) Are the differences of the mean performances of the children among the four levels of conservation of numerousness the same across the problems with a described transformation and problems without a described transformation?

9) Are the differences of the mean performances of the children among the four levels of conservation of numerousness the same across the problems involving the three levels of visual aids?

10) Are the differences of the mean performances of the children among the three IQ groups the same across the problems with a described transformation and problems without a described transformation?

11) Are the differences of the mean performances of the children among the three IQ groups the same across problems involving the three levels of visual aids?

12) Are the differences of the mean performances of the children the same for the problems describing a transformation and problems not describing a transformation across the three levels of visual aids?

13) Are the differences of the mean performances of children at the four levels of conservation of numerousness the same across the three IQ levels?

14) Are the differences of the mean performances of the children among the four levels of conservation of numerousness and three IQ groups the same across the problems describing a transformation and problems not describing a transformation?

15) Are the differences of the mean performances of the children among the four levels of conservation of numerousness and three IQ groups the same across the problems involving the three levels of visual aids?

16) Is the mean performance of the children in the four levels of conservation of numerousness different on a test of number facts ?

17) Is the mean performance of the children in the three IQ groups different on a test of number facts ?

18) Is there a significant correlation between children's ability to solve addition problems and their knowledge of arithmetic facts ?

THE PILOT STUDY OF ADDITION PROBLEMS

A pilot study was conducted the first part of December, 1965, using second graders, in order to obtain further empirical evidence on some of the above questions.[80] Secondgraders were used since they had been through an entire first grade curriculum and, at the time the pilot study was conducted, had been through almost half of a second grade curriculum. A study of problems with a subtractive structure was concurrently conducted, so in order to obtain optimal empirical information from the pilot study for both studies, addition problems and subtraction problems were used in the pilot study.

The subtraction test with a physical transformation was easier than the verbal addition test with a transformation and the pictorial addition test with no transformation was easier than the verbal subtraction test with no transformation. A surprising result is that the children in Level 4 showed up quite well on the subtraction test with a transformation and also did better on the verbal addition test with a transformation than on the addition and subtraction tests without transformation. Another significant result is that those children in Level 4 had a lower score for the four tests than did those children in the upper three levels. In the first three levels, the verbal subtraction test with no transformation was more difficult than the other three tests. Since a verbal addition test without a transformation was not administered, what this can be attributed to is not yet entirely clear, but since the verbal addition test without a transformation should be still more difficult (since we have indicated that problems with aids are easier to solve than problems without aids, holding transformation constant), there is evidence that a transformation does make problems easier to solve.

IV

DESIGN OF THE STUDY

SUBJECTS

The population for the study consisted of 2,166 first grade children from the Unified School District of Racine, Wisconsin. A sample of 341 children was randomly selected from this population. The order in which the sample was selected was recorded and used as the study progressed.

In December of 1965, each of the 2,166 children took the Kuhlmann-Anderson Intelligence Test, Form A, which was administered by the first-grade teachers of the school district. All the tests were graded by one grader. The frequencies of the IQ's from 78 to 140 are given in Table 4.

The IQ point of 100 corresponds to a cumulative frequency of 762 which is 40 more than one-third of total cumulative frequency of 2,166. Since the IQ point of 99 corresponds to a cumulative frequency of 701, which is less than one-third of the total cumulative frequency, the IQ range of 78-100 was selected as the first IQ range. Similarly 101-113 was selected as the second IQ range, and 114-140 was selected as the third IQ range. The average IQ of the population was 107.18 and the standard deviation 13.78.

MATERIALS AND PROCEDURES

After the selection of the ordered random sample of 341 children from the population, testing was begun on March 8, 1966. One trained experimenter did all testing. Each child was tested individually on three successive tests which took approximately 20 minutes.

The first test consisted of the conservation of numerousness test described in Chapter III from which the children were assigned to one of four levels. Since, as noted earlier, the study being conducted concurrently made it possible for two children to be taking the pretest together, it was decided not to randomly assign the three pretests to the children. The

Table 4

Frequencies of IQ from 78 to 140 for 2,166 First Grade Children

IQ	78	79	80	81	82	83	84	85	86
Frequency	10	13	12	7	15	18	20	23	25
IQ	87	88	89	90	91	92	93	94	95
Frequency	21	23	38	30	29	43	38	49	40
IQ	96	97	98	99	100	101	102	103	104
Frequency	65	59	48	75	61	57	60	50	46
IQ	105	106	107	108	109	110	111	112	113
Frequency	48	37	58	55	59	42	66	65	47
IQ	114	115	116	117	118	119	120	121	122
Frequency	47	64	30	46	65	46	18	38	48
IQ	123	124	125	126	127	128	129	130	131
Frequency	20	33	31	19	22	38	7	31	14
IQ	132	133	134	135	136	137	138	139	140
Frequency	11	17	10	6	14	5	15	10	9

pilot study of the test on conservation of numerousness indicated that Test 1 was easier than either of the other two tests, so it was administered first, then Tests 2 and 3, in that order. It turned out that Test 1 was still much easier than either 2 or 3. (See Chapter V.)

The second test consisted of eighteen addition problems. (See the Appendix.) Nine of these problems involved transformation and nine involved no transformation. Six had accompanying physical aids, six accompanying pictorial aids, and six no accompanying aids. Of each set of six problems, three involved transformation and three no transformation. Addition combinations that summed to from 5 through 8 were randomly assigned to the problems. No combination involving 1 as an addend was used. All the words used in the problems to describe the objects were 1) included as part of the curriculum, or 2) on "Clarence R. Stone's Revision of the Dale List of 769 Easy words," or on combinations of 1 and 2. The 18 problems were randomly assigned to each child and were typed on cards. The experimenter and the child sat opposite each other at a portable card table. The experimenter selected the appropriate card and, depending on the type of problem, she would 1) (a) if it was a problem involving a transformation put the first group of objects in front of the child and then read the problem to the child putting the second group of objects in front of the child as she read the phrase describing the transformation, or (b) if it was a problem involving no transformation, put all the objects in front of the child and then read the problem to the child; 2) put the picture in front of the child and read the problem to the child; or 3) just read the problem to the child. In all cases, no time limit was placed on the child and the problem was reread if the child wished. All questions the child may have asked were also answered if they did not involve the correct response. Moreover, if a child did not respond within a reasonable length of time (determined by the experimenter), the problem was read again. If the child still did not respond, the experiment was terminated for that child. Fortunately such trouble was encountered only in one or two cases.

The third test was a test of addition combinations with sums of 5, 6, 7 or 8. Again, no time limit was placed on the child to finish this test.

THE SAMPLING PROCEDURE

Originally, eight hundred children were selected as an ordered random sample from the total first grade population. Since two studies were in progress concurrently, 400 of these children were assigned to each study, with the children from each school randomly partitioned into two groups so that the two experimenters would have approximately an equal number of children for each school. Each subgroup of 400 still constituted an ordered random sample. The children in the sample from each school were tested. Any child in the sample who was absent on the day of testing for a particular school was not tested unless the experimenters returned to that same school the next day. At the end of the testing (April 12), an assignment of a subset of the 339 children tested in the IQ range of 78-140 to the twelve groups was made from the ordered sample by progressing through the sample in order assigning children to the appropriate groups until, say, group n was filled. Those children left in the unassigned portion of the sample who also were in group n were automatically discarded. This procedure was followed until all 12 groups were filled to 9 subjects. Additional testing was then performed on May 11 and 12 with children who were randomly selected from a randomly selected school until two more children were obtained for the lowest group (Level 4, IQ range 78-100) thereby increasing that group to 11. All other groups were increased to 11 by reverting back to the original remaining ordered sample and proceeding as before. In effect, the last sample also constituted an ordered sample which was considered an extension of the first sample. However, no more than two children described needed to be used for Sample 2.

THE EXPERIMENTAL DESIGN

Table 5 is a diagram of the design. N is the number of subjects per group (11) and g is the number of groups (12). Winer[81] outlines a repeated measures design whereby the main effects of 1) visual aids and 2) a transformation vs. no transformation and 3) levels and 4) IQ and all possible interactions are tested which will answer Questions 4-15 in the statement of the problem. The analysis of variance (ANOVA) table for this design is given in Table 6,[82] in which all factors are assumed to be fixed, as is the case in this study.

Table 5

Outline of the Design

Group	Individual	Tests 1	2	...	6	
1	1	X_{111}	X_{121}	...	X_{161}	
	2	X_{211}	X_{221}	...	X_{261}	
	⋮					
	n	X_{n11}	X_{n21}	...	X_{n61}	
	Means:	$X_{.11}$	$X_{.21}$	...	$X_{.61}$	$X_{..1}$
	⋮					
g	1	X_{11g}	X_{12g}	...	X_{16g}	
	2	X_{21g}	X_{22g}	...	X_{26g}	
	⋮					
	n	X_{n1g}	X_{n2g}	...	X_{n6g}	
	Means:	$X_{.1g}$	$X_{.2g}$	...	$X_{.6g}$	$X_{..g}$
Means: All Groups:		$X_{.1.}$	$X_{.2.}$	...	$X_{.6.}$	$X_{...}$

Table 6

ANOVA Table

Source of Variation	d. f.
<u>Between Subjects</u>	<u>npq-1</u>
A (Levels)	p-1
B (IQ)	q-1
AB	(p-1)(q-1)
Subj. w. groups	pq(n-1)
<u>Within Subjects</u>	<u>npq(rs-1)</u>
C (Visual aids)	r-1
AC	(p-1)(r-1)
BC	(q-1)(r-1)
ABC	(p-1)(q-1)(r-1)
C × subj. w. groups	pq(n-1)(r-1)
D (Action, no action)	s-1
AD	(p-1)(s-1)
BD	(q-1)(s-1)
ABD	(p-1)(q-1)(s-1)
D × subj. w. groups	pq(n-1)(s-1)
CD	(r-1)(s-1)
ACD	(p-1)(r-1)(s-1)
BCD	(q-1)(r-1)(s-1)
ABCD	(p-1)(q-1)(r-1)(s-1)
CD × subj. w. groups	pq(n-1)(r-1)(s-1)

Note: p = levels of factor A
q = levels of factor B
r = levels of factor C
s = levels of factor D

Greenhouse and Geisser[83] outline a procedure for developing F tests to detect: a) differences in the means of the six tests, b) differences in the means of the twelve groups, and c) differences in the test profile among the groups. These three tests will answer Questions 1, 2 and 3 in the statement of the problem. The ANOVA table for their procedure is given in Table 7.[84] In this table, the groups factor is subdivided into factors A and B and the interaction AB in Table 6. The subjects within groups is the same in both tables and is an error term in both analyses. The test factor of Table 7 is subdivided into factors C and D and the interaction CD of Table 6. The groups × tests interaction of Table 7 is subdivided into the interactions AC, BC, ABC, AD, BD, ACD, BCD, and ABCD of Table 6. The subjects within groups × tests interaction of Table 7 is subdivided into the C × subjects within groups, D × subjects within groups, and CD × subjects within groups interactions of Table 6.

The conservative tests outlined by Greenhouse and Geisser[85] will be used when testing for any within-subject variation. For Table 7 this amounts to entering the F table with 1 and N-g degrees of freedom and g-1 and N-g degrees of freedom instead of (V-1) and (V-1)(N-g) and (V-1)(g-1) and (N-g) respectively. For Table 6, the degrees of freedom used to test for variation within subjects is summarized in Table 8.

A two-way analysis of variance will be used to detect any possible differences in the performance of the children on the test of number facts in the four levels and three IQ groups which will answer Questions 16 and 17 in the statement of the problem. Correlation coefficients will also be calculated between total scores on the problem-solving test and total scores on the number-combination test to answer Question 18 of the statement of the problem.

Table 8

Degrees of Freedom for Conservative Tests

Source of Variation	d.f.
Within Subjects	
C	1
AC	p-1
BC	q-1
ABC	(p-1)(q-1)
C × subj. w. groups	pq(n-1)
D	1
AD	(p-1)
BD	(q-1)
ABD	(p-1)(q-1)
D × subj. w. groups	pq(n-1)
CD	1
ACD	p-1
BCD	q-1
ABCD	(p-1)(q-1)
CD × subj. w. groups	pq(n-1)

Table 7

ANOVA Table

Source of Variation	d.f.	SS	
Groups	g-1	Q_2	$F_2 = \frac{(N-g)}{(g-1)} \cdot \frac{Q_2}{Q_3}$
Sub. w. groups	N-g	Q_3	
Tests	V-1	Q_1	$F_1 = (N-g) \cdot \frac{Q_1}{Q_5}$
Groups × tests	(V-1)(g-1)	Q_4	$F_3 = \frac{(N-g)}{(g-1)} \cdot \frac{Q_4}{Q_5}$
Subj. w. groups × tests	(V-1)(N-g)	Q_5	

V

R' JLTS OF THE PRETEST AND OF THE RELIABILITY STUDIES

In this chapter, reliability studies of (1) the tests of conservation of numerousness, (2) the test of arithmetic problems, and (3) the subtests of the test of arithmetic problems that are of interest for the study will be discussed. The purpose for conducting these reliability studies is that when interpreting the results of a statistical analysis such as will be performed on the scores children received on the test of problem solving, it is essential to know the reliabilities associated with the tests on which the statistical analysis is based.

THE FOUR LEVELS OF CONSERVATION OF NUMEROUSNESS

The following discussion is divided into three sections: (1) the performance of the children on the pretest of conservation of numerousness which will be related to the pilot study and to a follow-up study, (2) reliability consideration of the pretest, and (3) the relation of the four levels to IQ.

The Performance on the Pretest of Conservation of Numerousness

Table 9 summarizes the frequency and proportion of the 341 children in the main study and the 100 children in the pilot study in each of the four levels. $\chi^2_3 = .13$ shows no significance in the departure of the proportions of the two groups considering the proportions for the main study as the expected values. The test then, functioned much the same way in the two samples even though there was a difference of at least five months between the two testings.

Disregarding guessing, as noted in Chapter III, in order for a child to score Item 4 of each test correctly, he had to make an extensive quantitative comparison. However, for Items 1, 2 and 3 of each test, a gross or intensive quantitative comparison would suffice. It seems, then, that in order for the definition of Level 4 to be meaningful, the frequencies of the first three items of each test should be much higher than the frequencies of the fourth item for each test. That this is in fact true for the 128 children in Level 4 is shown by Table 10.

Table 9

Frequency of Children in the Four Levels: Main Study and Pilot Study

Level	Main Study Freq.	Main Study Prop.	Pilot Study Freq.	Pilot Study Prop.
1	60	.18	8	.08
2	69	.20	20	.20
3	84	.25	31	.31
4	128	.38	23	.23

Table 10

Frequency of Children at Level 4 on Pretest
N = 128

Test	Item 1	Item 2	Item 3	Item 4
1	114	123	109	5
2	87	126	72	1
3	108	120	85	7

Based on a chance response, the frequencies under Item 4 could be expected to be as high as 43. They are much lower than this, indicating that generally the children were not guessing on this item but were possibly basing their judgments on gross quantitative comparisons. However, the possibility still exists for a small proportion of the children at Level 4 to be capable of making extensive quantitative comparisons. (No larger than about 1/30 as can be seen from Table 11; that is, a four-tuple having a 1 for the last entry and at least one 0 in the first three entries.)

Table 11

Frequency of Children's Response Patterns by Levels

	(1111)	(1110)	(1101)	(1011)	(0111)	(1100)	(1010)	(1001)	(0011)	(0110)	(0101)	(1000)	(0100)	(0010)	(0001)	(0000)
Level 1	180	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Level 2	128	50	6	4	1	1	1	-	-	2	2	-	2	-	-	-
Level 3	84	125	7	3	5	9	2	1	-	10	1	-	4	-	-	1
Level 4	-	225	3	1	4	69	8	-	-	28	2	3	38	-	2	1
Total	402	400	16	8	10	79	11	1	-	40	5	3	44	-	2	2

Key: 1 = Correct Response; 0 = Incorrect Response; (Item 1, Item 2, Item 3, Item 4)

Table 12

Frequencies of 20 Children on the Pretest: Non-Random Order

Test	Item 1	Item 2	Item 3	Item 4
1	19	18	19	15
2	16	18	15	10
3	17	18	15	8

Table 13

Frequencies of 20 Children on the Pretest: Random Order

Test	Item 1	Item 2	Item 3	Item 4
1	20	20	20	15
2	19	19	17	10
3	19	19	18	10

Table 14

Frequencies of Children at Level 3 on Pretest
N = 84

Test	Item 1	Item 2	Item 3	Item 4
1	83	82	81	69
2	66	84	69	20
3	82	79	79	12

Table 15

Frequency of Tests Entirely Correct, Level 2

Tests Entirely Correct	1, 2, 3	1, 2	1, 3	2, 3	1	2	3	None
Frequency	60	36	26	7	66	11	7	128

Table 16

Frequencies of Children at Level 2 on Pretest
N = 69

Test	Item 1	Item 2	Item 3	Item 4
1	69	68	67	63
2	64	69	64	47
3	67	65	65	41

Table 17

Frequencies of All Children on the Pretest
N = 341

Test	Item 1	Item 2	Item 3	Item 4
1	326	333	317	197
2	227	339	265	128
3	317	324	289	120

The possibility of a response bias exists since the items were taken in the order 1-2-3-4. However, if a classification at a level was the result of a response bias, then no differences should exist between levels. This is not the case, as will be shown later. Too, if the children had been responding on the basis of extensive quantitative comparisons, then the type of response given on the first three items would have been irrelevant. However, it is still true that response bias could exist. In order to gain more information about this question, a further study was conducted using a sample of 20 children who had just completed the first grade and were attending summer school. Ten of the children were given the tests just as they were given to the 341 children in the main study and the remaining 10 were given the tests and the items of each test in a random order all on the same day. Three days later, the 10 children who took the tests in a non-random order took it again, but this time in a random order, and the 10 children who took it in a random order took it again, but this time in a non-random order. Table 12 gives the frequency distribution for the non-random administration of the test and Table 13 gives the frequency distribution for the random administration. $\chi^2_{11} = 2.60$ (not significant at the .05 level) shows that the frequencies of the two tables are not different except for chance fluctuation.

Tables 14 and 15 show that Test 1 was much easier than Test 2 or 3 for the 84 children in the main study in Level 3. A plausible explanation has been already given in Chapter III for this phenomena since the data in Table 15 is quite consistent with the data in Table 3.

The children in the main study in Level 2 usually scored correctly on Tests 1 and 2 or 1 and 3 as is shown by Table 15. Table 16 shows the frequencies of these children on all items, and Table 17 shows the total frequency for the 341 children on all the items of all the tests.

Reliability Consideration of the Pretest

On any one test, a child could score 0, 1, 2, 3, or 4. Using these scores, correlations were calculated between Tests 1 and 2, 1 and 3, and 2 and 3. The results, given in Table 18, may be considered as comparable form reliability coefficients.[86] The coefficients are significantly different from a 0 correlation but, since they are modest correlations, they support the necessity of having more than one test of conservation of numerousness.

Table 18

Correlations Between Total Scores on the Three Subtests of the Pretest

	Test 1	Test 2	Test 3
Test 1	1	.37**	.40**
Test 2		1	.46**
Test 3			1

** $p < .01$

If the 12 items of all three tests are considered as one test, the concept of an internal consistency reliability coefficient becomes meaningful.[87] Table 19 gives the inter-item correlation coefficients of the 12 items where variables 1-4 are the items of Test 1, 5-8 are the items of Test 2 and 9-12 are the items of Test 3, all similarly ordered. Pearson product-moment correlation coefficients were used ($\emptyset$ coefficients). A computer program was used to obtain the table.[88]

Using the Spearman-Brown Prophecy formula $\left(\frac{k \cdot \bar{r}}{1 + (k-1)\bar{r}}\right.$, where k = number of items and $\bar{r}$ = average inter-item correlation of the k items) for the internal consistency reliability coefficient, a respectable coefficient of .69 is obtained.[89] Considering just Items 4, 8, and 12 (those requiring the response "the same number") an internal-consistency coefficient of .72 is obtained, quite respectable for a 3-item subtest. The inter-item correlations of these three items are given in Table 20. For a 6-item test of the same kind of items, a reliability estimate of .84 is given by the Spearman-Brown Step-up formula[90] $\left(r_n = \frac{n \cdot r}{1 + (n-1)r}\right.$, where r is the original reliability, and r_n is the estimated reliability of a test n times as long).

Considering the first three items of each test (variables 1, 2, 3, 5, 6, 7, 9, 10 and 11), an internal-consistency reliability coefficient of .61 is obtained. An inspection of Table 17 indicates that, because of the large frequencies, many small insignificant and significant inter-item correlation coefficients should be obtained. This is in fact the case as 26 of the possible 66 inter-item correlations do not differ significantly from 0 and hence contribute heavily to the unreliability of the total test. Eighteen of these insignificant correlations involve correlations between the first three items of each test, of which there are only 36. As

Table 19

Inter-Item Correlation Matrix of the Pretest

	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	Var 8	Var 9	Var 10	Var 11	Var 12
Var 1	1.00	.06	.16**	.19**	.15**	-.02	.13*	.13*	.28**	.08	.23**	.10
Var 2		1.00	.18**	.06	-.02	.24**	.06	.04	-.04	.14**	-.07	.07
Var 3			1.00	.21**	.31**	-.02	.29**	.14**	.20**	-.01	.36**	.11*
Var 4				1.00	.21**	.09	.37**	.47**	.22**	.02	.30**	.41**
Var 5					1.00	-.04	.36**	.23**	.17**	.17**	.23**	.17**
Var 6						1.00	-.04	.06	-.02	.16**	-.03	.06
Var 7							1.00	.28**	.31**	.20**	.52**	.16**
Var 8								1.00	.16**	.01	.24**	.49**
Var 9									1.00	.10	.37**	.05
Var 10										1.00	.09	-.09
Var 11											1.00	.14**
Var 12												1.00

*$p < .05$

**$p < .01$

noted, the inter-item correlation of the last items of the test are quite substantial and result in a quite reliable subtest.

In order to gain more information on the stability of the levels, a correlation coefficient was computed between the levels the 20 children in the follow-up study were placed in on the two days of testing. (See Table 21.) Four of the five children who changed levels improved on the second day of testing. Since the tests were given only two days apart, some improvement should be expected due to the familiarity of the testing situation on the second day. The correlation coefficient of .78 obtained indicates, however, good stability of the levels.

The Relation of the Four Levels of Conservation of Numerousness and IQ

The frequencies of levels by IQ are given in the Appendix. Collapasing this frequency table on the IQ variable into the three ranges defined for the study Table 22 is obtained, from which $\chi^2_6 = 61.15$ is significant beyond the .001 level of significance. The mean IQ of the total sample as well as those of the four levels are given along with the standard deviation of the sample in Table 23.

As noted earlier, 11 children from each of the cells in Table 22 were randomly selected for the remainder of the study.

Table 20

Inter-Item Correlation of Items 4, 8, and 12 of the Pretest

Items	4	8	12
4	1	.47	.41
8	-	1	.49
12	-	-	1

Table 21

Levels Achieved by 20 Children on Two Different Days

Day	Subject 1	2	3	4	5	6	7	8	9	10
Tues.*	1	1	2	4	3	2	2	3	2	4
Fri.	1	1	2	2	2	2	2	3	2	4

Day	Subject 11	12	13	14	15	16	17	18	19	20
Tues.	1	1	4	3	2	3	2	4	4	4
Fri.*	1	1	3	2	4	3	2	4	4	4

*Random Order

Table 22

Frequency Table of Levels by IQ

IQ	Level 1	2	3	4	Total
98–100	11	15	19	69	114
101–113	11	18	34	37	100
114–140	38	36	31	22	127
Total	60	69	84	128	341

Table 23

Mean IQ's for Levels and Total Sample and Standard Deviation of Total Sample

	Level 1	2	3	4	Total Sample
Mean IQ	116.23	112.87	109.76	98.20	108.21
Std. Dev.	-	-	-	-	14.33

RELIABILITY STUDIES OF THE TEST ON PROBLEM SOLVING

The following reliability studies are based on the performance of the sample of 132 children on the test of addition problems. There will be a total of 12 reliability coefficients reported, one for the total test and the others for subtests as well as an item analysis using the total test as an internal criterion.

Total Test

On the problem solving test, it was possible to obtain a total score from 0 to and including 18. Table 24 gives the frequency distribution of these total scores.

For each item, four statistics are computed: (1) difficulty, (2) item-criterion correlation, (3) x_{50}, and (4) beta. Underlying these four statistics is the concept of an item characteristic curve, which "is a smooth curve fitted to the proportion of persons at each criterion score level who made the particular response being studied." [91] In order to utilize the parameters of the normal curve (μ, σ), the assumption that the item characteristic curve has the form of the integrated normal function (normal ogive) must be made. [92] Once this assumption is made, the definition of x_{50} and β may then be given.

> The parameters of the item characteristic curve which specify the normal ogive fitted to the item response data are the following.
>
> x_{50}, the criterion score at which the probability of correct response is .5. The parameter is expressed in units of the criterion variable standard deviation.
>
> β, a measure of the steepness of the item characteristic curve which specified the capability of the item to discriminate between the individuals possessing various amounts of the criterion ability. This parameter is the reciprocal of the standard deviation of the fitted normal ogive.[93]

β may also be thought of as "the slope of the item characteristic curve at x_{50}."[94]

The difficulty of an item "corresponds to the area under the item characteristic curve and hence is a function of both x_{50} and β."[95]

The correlation of an item with a criterion score, such as total test score, may be computed by using point biserial correlation, which assumes the ability underlying the responses on each item is continuous. The formula for point biserial correlation is $r_b = \frac{\bar{x}_1 - \bar{x}}{s_x} \frac{(P)}{(Z)}$,[96] where $\bar{x}_1$ is the mean score of all persons answering the item correctly; $\bar{x}$ is the mean of the sample; s_x is the standard deviation of the sample; P is the proportion of persons answering the item correctly and Z is the ordinate of the normal curve at the deviate which divides the area of the unit normal curve into P and 1-P. A relationship between r_b and β is given

Table 24

Frequency Distribution of Total Scores on Problem Solving Test

Total Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Frequency	1	1	-	1	-	-	-	3	1	3	7	7	5	13	6	12	24	26	22

Mean: 14.56
Std. Dev.: 3.49

by $\beta = \frac{r_b}{\sqrt{1-r_b}}$,[97] which is utilized in the computation of β.

The Hoyt reliability coefficient, an internal consistency reliability coefficient,[98] is given by the formula $r_h = 1 - \frac{MS_w}{MS_{ind}}$,[99] where MS_w and MS_{ind} are two mean squares in a two-way analysis of variance in which one of the factors is the subjects and the other is the test items.[100] The above procedures were used in all item analyses that follow, which were performed by a computer program.[101]

Table 25 contains the analysis of variance used to compute the Hoyt reliability coefficient for the problem solving test and the reliability coefficient.

An inspection of the table shows a significant difference of the individuals on the test. An inspection of Table 24 shows the scores distributed from 0 to 18 of which the significant difference is a reflection. A more detailed analysis of the individuals will be given later when the performance of the children in the 12 groups—the 4 levels and 3 IQ groups—are studied. The significant *F* ratio associated with the items indicates that some were more difficult than others which is seen in the item analysis that follows. The reliability coefficient of .83 indicates that the test is quite reliable and compares favorably with commercially produced tests,[102] of which one reports internal consistency reliability coefficients from .78 to .93. In this test battery, the arithmetic reasoning test of 45 items has a reliability coefficient of .84, and the arithmetic fundamentals, addition and subtraction, a test of 45 items, has a reliability coefficient of .93. For a test three times as long as the addition problem test (a test with 48 items) the estimated reliability is .93 as calculated by the Spearman-Brown Step-Up Formula.

Table 25

ANOVA Table for Hoyt Reliability of Total Test

Source of Variation	d.f.	S.S.	M.S.	F	H.R.
Ind.	131	89.14	.78	6.06**	.83
Items	17	28.13	1.65	14.74**	
Error	2227	249.98	.11		
Total	2375	367.25			

**p < .01

Table 26 contains the item analysis of the test. The criterion variable is the total test score on which the statistics are based.

Due to the fact that the test was constructed

Table 26

Item Analysis of Problem Solving Test

Item	Frequency of correct responses	Difficulty	r_b	x_{50}	Beta
1	122	.92	.77	-1.86	1.21
2	113	.86	.86	-1.24	1.66
3	122	.92	.65	-2.21	.85
4	117	.89	.95	-1.27	3.08
5	110	.83	.93	-1.04	2.56
6	100	.76	.83	- .84	1.50
7	125	.95	.76	-2.12	1.18
8	111	.84	.55	-1.80	.66
9	103	.78	.68	-1.13	.94
10	111	.84	.97	-1.02	4.21
11	103	.78	.93	- .83	2.59
12	118	.89	1.09	-1.15	—
13	105	.80	.69	-1.19	.95
14	101	.77	.57	-1.27	.69
15	111	.84	.72	-1.38	1.04
16	97	.73	.82	- .76	1.42
17	60	.45	.57	.20	.69
18	93	.70	.70	- .77	.97

to test the problem solving abilities of first grade children who had been through three-fourths of a first-grade arithmetic curriculum, it is to be expected the test should be quite easy for those children tested. This expectation is verified by the item analysis in that all items except 17 have x_{50}'s below the mean of the criterion test. Moreover, generally high item-criterion correlations are present, with large difficulty indices. However, the items generally are good discriminators, with only 3 items possessing a β below .85 (a slope of about 40°). The smallest slope observed corresponds to an angle of about 33-1/2°. The mean of the 18 items for all 132 children was 14.76 and the standard deviation 3.49. One standard deviation below the mean corresponds to a score of about 11, and two standard deviations below the mean corresponds to a score of about 8. Since a large number of items function between 8 and 11, these items are discriminators for approximately 6 to 24 children. Items which function from one standard deviation below the mean to the mean are discriminators for about 24 to 50 children. The relationship of these lower scores to IQ and conservation of numerousness will be made clear later. The last 6 items were more difficult than the first 12 indicating that problems with no accompanying aids were more difficult than the problems with accompanying aids. No other apparent differences in difficulty are discernable at this time except for a possible difference in 13, 14, 15 as opposed to 16, 17, and 18 (verbal problems with a transformation as opposed to verbal problems without a transformation) and must await further analysis.

Subtests

Detailed information on the remaining reliability studies is available in the original report.[103] Reliability coefficients are given in Table 27.

SUMMARY

The pretest of conservation of numerousness partitioned two independent samples much the same way even though five months separated the two testings. Randomizing the items of the pretest and the three tests does not seem to affect the way in which children respond to the items thus eliminating the possibility of response bias influencing the level in which children are placed. Those children who were placed in Level 4 scored very low on Item 4 of each test of the pretest which was very consistent with the definition of Level 4, which was based on the type of quantitative comparisons of which children are capable. Test 1 turned out to be much easier than either 2 or 3 which is consistent with the results of the pilot study of conservation of numerousness.

A short time-interval test-retest reliability coefficient of .78 was obtained for the pretest. An internal-consistency reliability coefficient of .69 also was obtained. Both are substantial and indicate that the results of the pretest may be interpreted with some confidence. IQ and the four levels are associated as determined by a significant χ^2.

To facilitate the interpretation of the results, reliability studies were conducted on the test

Table 27

Internal-Consistency Reliability Coefficients of Problem Solving Subtests

Subtest	Number of Items	Reliability
1. Physical Aids; Transformation	3	.40
2. Physical Aids; No Transformation	3	.60
3. Pictorial Aids; Transformation	3	.47
4. Pictorial Aids; No Transformation	3	.77
5. No Aids; Transformation	3	.57
6. No Aids; No Transformation	3	.37
7. Physical Aids	6	.64
8. Pictorial Aids	6	.69
9. No Aids	6	.65
10. Transformation	9	.65
11. No Transformation	9	.81

of addition problems and 11 subtests of the test of addition problems. Item analyses were included in the reliability studies. The tests (or subtests) and reliabilities associated with them are given in Table 27.

Difficulties of interpretation may arise in analyses that involve the six three-item subtests (1-6 in Table 27) due to relatively low reliabilities associated with three of them (1, 3, and 6). Difficulties are not as prevalent for 1 as for 3 and 6 since the items in 1 were of very low difficulty and were not statistically different. But in the case of 3 and 6, the item difficulties departed significantly one from the other, with those of 6 being much more difficult than those of 3.

Considering the total problem solving test as the criterion, the items all functioned below the mean (had x_{50}'s below the mean) except one (number 17) which functioned above the mean in all item analyses. However, the items generally were good discriminators at the x_{50} points with the exception of three, which had β's below .69. The item analyses performed using the subtests as the criterion revealed that the two subtests denoted by 2 and 4 in Table 28 were quite highly related in that when Subtest 2 was used as a criterion, those items in Subtest 4 possessed high item-criterion correlations, good β's (those that were non-zero) and low x_{50}'s; and vice versa. No other pair of subtests in 1-6 of Table 27 displayed this high degree of relationship. However, when Subtest 7 was used as a criterion, the items of Subtest 4 possessed high item-criterion coefficients, low x_{50}'s and good β's, which is not surprising. Also, when Subtest 8 was used as the criterion, Subtest 2 possessed high item-criterion correlations, low x_{50}'s and good β's which again is not surprising. No subtest displayed a high relationship to subtests 5, 6, or 9, nor to 1, 3, 10, or 11.

VI

RESULTS OF THE STUDY

THE PERFORMANCE OF THE CHILDREN IN THE TWELVE GROUPS ON THE SIX PROBLEM TYPES

As noted earlier, the sample of first graders was divided into 12 groups of 11 children each on the basis of level of conservation of numerousness and IQ classification. Due to the large ranges of the IQ classifications (78-100, 101-113, 114-140, 78-140) it could be entirely possible for the mean IQ's of the children among different levels within a given IQ classification to differ significantly. With this in mind, four one-way analyses of variances were performed, one within each IQ classification, using levels as the main effect. Table 28 gives the mean IQ's of the children in each IQ classification and Tables 29, 30, 31, and 32 are the ANOVA tables. The _F_ ratios in the four analyses of variance tables are all less than 1, indicating no statistical differences in the mean IQ's among the four levels within each IQ classification given in Table 28. Any differences, then, that might occur between the mean performances of the children in the four levels are not due to differences in IQ.

Each child of each group of 11 children took all of the problems of each of the six tests. The design was outlined in Chapter IV to detect any possible differences in the mean performance of the children among the 12 groups, any possible differences in the means of the six tests and any possible differences in the group profiles. The analysis of variance is outlined in Table 33.

The _F_ ratio of 1.04 shows no statistically significant interaction of groups and problem types (no significant difference in the group profiles). Table 34 gives the group profiles (means) for the six problem types.

The _F_ ratio of 2.98 for the 12 groups was significant beyond the .01 level of significance indicating that the mean performances (Table 34) of the children among the 12 groups were statistically different.

Table 28

Mean IQ's of Children Among the Four Levels Across Four IQ Classifications

Level	IQ 78-100	101-113	114-140	78-140
1	93.18	108.09	124.36	108.55
2	94.18	107.91	121.73	107.94
3	94.18	106.82	122.36	107.79
4	91.73	108.73	119.82	106.76

Table 29

ANOVA for IQ Range 114-140 Across Four Levels

Source of variation	d. f.	MS	F
Between Levels	3	38.629	< 1
Within Levels	40	47.523	
Total	43		

Table 30

ANOVA for IQ Range 101-113 Across Four Levels

Source of variation	d. f.	MS	F
Between Levels	3	6.931	< 1
Within Levels	40	15.091	
Total	43		

Since the interaction of groups and problem types is not significant, the Newman-Keuls method of testing the difference between all possible pairs of means can be used.[104] A description of the test will be embedded in the procedure that follows.

In Table 34 the means of all groups are rank-ordered from low to high. The differences

Table 31

ANOVA for IQ Range 78-100 Across Four Levels

Source of variation	d. f.	MS	F
Between Levels	3	14.818	< 1
Within Levels	40	39.471	
Total	43		

Table 32

ANOVA for IQ Range 78-140 Across Four Levels

Source of variation	d. f.	MS	F
Between Levels	3	18.202	< 1
Within Levels	128	161.458	
Total	131		

Table 33

ANOVA for the Six Problem Types and Twelve Groups

Source of variation	d. f.	SS	MS	F
Groups	11	57.662	5.242	2.98**
Subj. w. Groups	120	211.091	1.759	
Problem Types	5	51.434	10.287	27.09***
Groups × Types	55	21.747	.395	1.04
Subj. w. Groups × Types	600	227.818	.380	
Total	791	569.753		

**p < .01
***p < .01, Conservative Test

Table 34

Means of the 12 Groups on the Six Problem Types

Group	Type 1	2	3	4	5	6	Group Means	Group Ranks
G_{11}	2.82	2.36	2.64	2.55	3.00	2.27	2.61	9
G_{21}	3.00	2.82	2.82	3.00	2.82	2.27	2.79	12
G_{31}	2.82	2.36	2.64	2.45	2.09	2.09	2.41	6
G_{41}	2.55	2.36	2.18	2.18	2.45	1.55	2.21	2
G_{12}	2.91	2.82	2.91	3.00	2.45	2.27	2.72	11
G_{22}	2.73	2.73	2.82	2.27	2.82	1.73	2.52	7
G_{32}	2.91	2.64	2.82	2.73	2.45	2.09	2.61	10
G_{42}	2.55	2.36	2.09	2.55	2.36	1.91	2.30	3
G_{13}	2.82	2.73	2.91	2.55	2.27	2.00	2.55	8
G_{23}	2.82	2.36	2.45	2.18	2.36	1.73	2.32	4
G_{33}	2.64	2.36	2.36	2.55	2.45	1.82	2.36	5
G_{43}	1.91	1.82	2.09	2.18	1.36	1.00	1.73	1
Problem type means	2.70	2.48	2.56	2.52	2.41	1.89		

Note: For group designation, first digit indicates level of conservation of numerousness (1–4); second digit indicates IQ group (1 = 114–140, 2 = 101–113, 3 = 78–100).

Table 35

Difference Between All Pairs of Means of the 12 Groups
(Groups listed in rank order)

	G_{41}	G_{42}	G_{23}	G_{33}	G_{31}	G_{22}	G_{13}	G_{11}	G_{32}	G_{12}	G_{21}
G_{43}	.485	.576	.591	.637	.682*	.788*	.819*	.879*	.879*	1.00*	1.061*
G_{41}	-	.091	.106	.152	.197	.303	.334	.394	.394	.515	.576
G_{42}	-	-	.015	.061	.106	.212	.243	.303	.303	.424	.485
G_{23}	-	-	-	.064	.091	.197	.228	.288	.288	.409	.470
G_{33}	-	-	-	-	.045	.151	.182	.242	.242	.363	.424
G_{31}	-	-	-	-	-	.106	.137	.197	.197	.318	.397
G_{22}	-	-	-	-	-	-	.031	.091	.091	.212	.273
G_{13}	-	-	-	-	-	-	-	.060	.060	.181	.242
G_{11}	-	-	-	-	-	-	-	-	.000	.121	.182
G_{32}	-	-	-	-	-	-	-	-	-	.121	.182
G_{12}	-	-	-	-	-	-	-	-	-	-	.061

*p < .05

Table 36

Critical Values for Newman-Keuls Test

r	2	3	4	5	6	7	8	9	10	11	12
$q_{.95}$ (r, 120)	2.80	3.36	3.69	3.92	4.10	4.24	4.36	4.48	4.56	4.64	4.72
$S_G\, q_{.95}$ (r, 120)	.457	.548	.602	.640	.669	.692	.712	.731	.744	.757	.770

between all possible pairs of means are computed and entered in Table 35.

To obtain the critical values beyond which the differences of the means are significant, the standard error of the mean,

$$S_G = \sqrt{\frac{MS_{\text{subj. w. groups}}}{n \cdot q}},$$

is computed and then $S_G \cdot Q_{1-\alpha}$ (r, d.f.), which is the critical value, is computed, where d.f. = degrees of freedom associated with S_G (120), r = the number of steps the two ordered means are apart in an ordered sequence, and $q_{1-\alpha}(r, \text{d.f.}) = \frac{\overline{M}_i - \overline{M}_j}{S_G}$, where $\overline{M}_i$ = mean of group i, and $\overline{M}_j$ = mean of group j, and q is the studentized range statistic. The above computations are summarized in Table 36. The two means G_{21} and G_{43} are 12 steps apart, so the difference must exceed .770, which it does, and is therefore significant. This procedure is continued for r = 11, 10, . . . until the first critical value, if any, is met which is not exceeded, in this case at r = 5. The process is then terminated for that row and performed for all subsequent rows.[105]

Using the above procedure for testing differences of group means, the difference of the means of the pairs (G_{43}, G_{21}), (G_{43}, G_{12}), (G_{43}, G_{32}), (G_{43}, G_{11}), (G_{43}, G_{13}), (G_{43}, G_{22}) and (G_{43}, G_{31}) are all significant as is shown in Table 35. The differences of the means of the four other groups and G_{43} were very close to being significant. This test indicates, then, that the mean performance of the children in the low IQ and low level classification is lower than all other groups and significantly lower than seven of them. The reliability of .83 reported in Chapter V for the total test indicates that the test was a reliable estimate of the children's ability to solve addition problems. Moreover, in the analysis of variance for each subtest reliability study, the main effect of individuals was always significant. In this analysis, groups of individuals have been identified which were contributing to that difference. An inspection of Table 22 of Chapter V shows that 69 of the 341 children in the sample were in group G_{43} (slightly over 20% of the sample).

It is interesting to note in Table 35 that within an IQ classification the means generally are lower for the lower levels. The same is true for IQ across levels. A more complete elucidation of these two factors considered separately will be deferred until later.

The F ratio of 27.09 which corresponds to problem types, shown in Table 34, is highly signigicant. The means are given in Table 34.

Again, the Newman-Keuls method of testing the differences between all possible pairs of means can be used. Table 37 gives the difference of all possible pairs and Table 38 contains the critical values.

The low reliabilities reported in Chapter V for Problem Types 1 (.40), 3 (.47), and 6 (.37) dictate that the results of this analysis must be interpreted with caution. However, the differences of the means of Problem Type 6 (verbal problems with no transformation) and all the other problem types far exceed the necessary critical values as is shown in Tables 37 and 38. In view of this fact and the fact that three of the six problem types (2, 4, and 5) had substantial reliability coefficients, the conclusion that the mean of Problem Type 6 was significantly lower than the means of the other five problem types is well supported. Moreover, the difference of the mean of the problems with accompanying physical aids and each of the Problem Types 2 (accompanying physical aids with no transformation), 4 (accompanying pictorial aids with no transformation), and 5 (verbal problems with a transformation) were significant.

THE PERFORMANCE OF THE CHILDREN IN THE FOUR LEVELS AND IN THE THREE IQ GROUPS

In this section, the 12 groups are subdivided into four levels and three IQ groups. Table 39 gives the analysis of variance for these two factors.

The interaction of levels and IQ is not significant, indicating there is no statistical difference of the IQ profiles among the four levels, which are given in Table 40.

Due to this lack of interaction, it is possible to again use the Newman-Keuls test of ordered means to contrast all possible differences of the means of the four levels and three IQ groups. Table 40 gives the means of the four levels, Table 41 gives the matrix of differences of the ordered means of the four levels, and Table 42 includes the critical values.

Table 43 gives the matrix of difference of the ordered means of the IQ groups reported in Table 40, and Table 44 includes the critical values.

Table 37

Difference Between All Possible Pairs of Means of Six Problem Types
(Problem types in rank order)

	6	5	2	4	3	1
6	-	.515*	.583*	.621*	.667*	.811*
5	-	-	.068*	.106	.152	.296*
2	-	-	-	.038	.084	.228*
4	-	-	-	-	.046	.190*
3	-	-	-	-	-	.144

*p < .05

Table 38

Critical Values for Neuman-Keuls Test of Ordered Means

r	2	3	4	5	6
$q_{.95}$ (r, 600)	2.77	3.31	3.63	3.86	4.03
$S_p\, q_{.95}$ (r, 600)	.149	.176	.195	.207	.216

$$S_p = \sqrt{\frac{MS_{\text{types} \times \text{subj. with groups}}}{n \cdot g}}$$

Table 39

ANOVA for the Four Levels and Three IQ Groups

Source of variation	d. f.	MS	F
Between subj.			
Levels (A)	3	11.450	6.51**
IQ (B)	2	7.085	4.03*
AB	6	1.524	< 1
Sub. w. groups	120	1.759	
Total	131		

$^*p < .05$
$^{**}p < .01$

Table 40

Means: Levels × IQ

Levels	IQ 114-140	IQ 101-113	IQ 78-100	Mean
1	2.61	2.73	2.54	2.63
2	2.79	2.52	2.32	2.54
3	2.41	2.61	2.36	2.46
4	2.21	2.30	1.73	2.08
Mean	2.50	2.54	2.24	

Table 41

Differences of Means of the Four Levels

Level	4	3	2	1
4	-	.379**	.459**	.545**
3	-	-	.080	.166
2	-	-	-	.086

$^{**}p < .01$

Table 42

Critical Values: Newman-Keuls Test (.01 Level)

r	2	3	4
$q_{.99}$ (r, 120)	3.70	4.20	4.50
$S_{\bar{S}}\, q_{.99}$ (r, 120)	.349	.396	.424

$$S_{\bar{S}} = \sqrt{\frac{MS_{\text{sub. w. groups}}}{\text{n. q. r. s.}}}$$

Table 43

Difference of the Means of the Three IQ Groups

IQ	78-100	114-140	101-113
78-100	-	.266*	.300*
114-140	-	-	.034

$^*p < .05$

Table 44

Critical Values of Newman-Keuls Test (.05 Level)

r	2	3
$q_{.95}$ (r, 120)	2.80	3.36
$S_I\, q_{.95}$ (r, 120)	.229	.274

$$S_I = \sqrt{\frac{MS_{\text{subj. w. groups}}}{\text{n. p. r. s.}}}$$

The overall F test to detect differences in the means of the four levels was significant at the .01 level of significance as is shown in Table 39. Moreover, Table 41 indicates that the mean performance of the children in Level 4 differed significantly (at the .01 level) from each of the other levels. However, no statistical differences exist between any pair of the means of Level 1, Level 2 and Level 3. As has been discussed earlier the proportion of children in Level 4 who were capable of making extensive quantitive comparisons was quite low. However, the probability was quite good (better than 2/3) that children in Level 3 had made extensive quantitative comparisons. The probability of children making extensive quantitative comparisons increased for each of Levels 2 and 1, and, in the case of Level 1, it was almost certain that a child had made an extensive quantitative comparison. Moreover, the differences in the mean performance of the children among the four levels are differences obtained in the mean performance of children on a problem solving test with an internal consistency reliability of .83. In summary, then, a random sample of the group of children (38 per cent of the sample) for which it was highly improbable that extensive quantitative comparisons had been made performed significantly lower on a test which reliably measured their ability to solve addition problems than children for which the probability was higher for making extensive quantitative comparisons. This result is not inconsistent with Piaget's postulate that conservation of something is a necessary condition for any mathematical understanding. (See page 3). The fact that the children at Level 4 did solve approximately 2 of every 3 problems does not, however, completely support Piaget's hypothesis if it is interpreted in the strictest sense. (Note the word <u>any</u> in the cited postulate.) This result is also not inconsistent with the correlation of .59 that Dodwell obtained between a test of conservation and a test covering the content of the first term of an arithmetic curriculum. Among differences that do exist, however, is the time of administering the pretest. In this study, the pretest was administered immediately before the problem test; and in Dodwell's study, the pretest was administered seven months before the test over the content was administered. Table 9 shows, however, that the pretest did function much the same in two different populations over a five-month time interval. This five-month interval should include much that is crucial for a good performance on the pretest (counting and correspondence, as well as age differential). However, no long-term test-retest reliability study has been made as of yet, but the above observation indicates considerable reliability over the five-month period.

With reference to the IQ variable, Cronbach notes that the Kuhlmann-Anderson Intelligence Test has been constructed pragmatically by "trying items and retaining those which correlate with such criteria as school success. . . . "[106] In this test nearly all of the subtests require adaptation to new situations but do depend on experience and require special abilities. No one specialized ability plays a large part in the score. However, verbal ability is important because the pupil must comprehend directions, but the test designers use simple vocabulary and introduce reading only in the later tests, and not in those for first grade. The subtests have a low correlation with each other, thus increasing the comprehensiveness of the test by bringing in many aspects of ability. The Kuhlmann-Anderson test measures substantially the same thing as the Binet.[107] The Binet scores are also strongly weighted with verbal abilities as the great majority of test items call for facility in using and understanding words. [108]

As shown by Table 43, the mean performance of the children in the IQ range of 78-100 was significantly lower than the mean performances of the children in the two other IQ ranges, which did not differ from each other. This lower mean performance certainly can be attributed, in part, to the fact that the Kuhlmann-Anderson test was constructed using items which correlated well with school success. Moreover, as noted, the Kuhlmann-Anderson test requires the pupils to comprehend directions, thus requiring a verbal ability. The tests on problem solving also required the children to comprehend the direction to find an answer for the question asked using the information they had been given by the experimenter.

The rank order of means given in Table 34 for all twelve groups indicates that the lowest mean was for the group of children in Level 4 and IQ range 78-100. This indicates that the mean performance of the children in the lowest combination of the two variables, about 20 per cent of the sample studied, was significantly lower than for many other combinations of the two variables, which is consistent with the above discussion.

THE EFFECT OF VISUAL AIDS

In this section, the effects of the six tests are partially subdivided into the aids variable of which there are three levels: Physical, pictorial, and no aids. The reliabilities of these three subtests were reported in Table 27 as .64, .69, and .65 respectively. Moreover, the profiles of the four levels, the three IQ groups, and the twelve groups are assessed with respect to the variable under study. Table 45 gives the analysis of variance pertaining to the above discussion.

The table shows a highly significant $\underline{F}$ ratio for the effect of aids in problem solving, but none of the interactions of aids and levels, aids and IQ and aids by levels and IQ were significant. The means of all the children with respect to the aids variable are given in Table 48. Tables 46 and 47 are the usual tables of the Newman-Keuls test. Instead of $q_{.95}$ (r, 240), $q_{.95}$ (r, 120) is used because linear interpolation between 120 degrees of freedom and ∞ degrees of freedom is impossible. Moreover, if significance is obtained

Table 45

ANOVA Table for Effect of Aids

Source of variation	d.f.	MS	F
Aids	2	15.187	37.91***
Aids × Levels	6	.126	< 1
Aids × IQ	4	.541	1.35
Aids × IQ × Levels	12	.589	1.47
Aids × Subj. w. Groups	240	.401	

***p < .01, Conservative Test

Table 46

Differences of Means of the Three Levels of Aids

	Verbal	Pictorial	Physical
Verbal	-	.386**	.439**
Pictorial	-	-	.053
Physical	-	-	-

**p < .01

Table 47

Critical Values: Newman-Keuls Test (.01 Level)

r	2	3
$q_{.99}$ (r, 120)	3.70	4.20
$S_A \cdot q_{.99}$ (r, 120)	.144	.164

$$S_A = \sqrt{\frac{MS_{\text{aids} \times \text{subj. w. group}}}{n. p. q. s.}}$$

using 120 degrees of freedom, then significance must be obtained using 240 degrees of freedom.

The above observations are consistent with those observed by Marilyn Zweng in the case of second graders solving subtraction problems. They are also consistent with Piaget's interpretation of concrete operations; that is, that concrete operations operate on objects and not yet on verbally expressed hypotheses. They do not, however, support Smedslund's speculation that early training should be given in the absence of perceptual support. It is true that the children have received a lot of training with perceptual support. They have also received training without such support, but perhaps not the same amount. There was no difference in the means of the problems with accompanying physical aids and accompanying pictorial aids. This could be a result of the training received (heavy emphasis on pictorial aids) or a result due to the fact that the use of physical aids is not superior to the use of pictorial aids as a training device.

There was no interaction of levels × aids as is shown by Table 48. The table reflects the fact that the mean performance of the children in Level 4 was significantly different from the mean performance of the three other levels and that the problems without aids (verbal) were significantly more difficult than the problems with accompanying aids. The children in Level 4 only made a mean score of approximately 59 per cent on the verbal problems; that is, they had an average of only about 3.5 of the 6 verbal problems. These same children were not highly successful on the problems with aids either, although more successful than on the verbal problems. They had an average score of only about 4.5 of the 6 problems with accompanying physical aids and an average of about 4.4 of the 6 problems with accompanying pictorial aids. The children in the top three levels scored just as well (on the basis of the mean score) on the verbal problems (those in Level 3 scored slightly lower) as the children in Level 4 did on the problems with accompanying physical aids.

Table 48

Interaction of Levels and Aids: Mean Scores

Levels	Aids		
	Physical	Pictorial	Verbal
1	2.74	2.76	2.39
2	2.74	2.59	2.29
3	2.62	2.59	2.17
4	2.26	2.21	1.77
Mean	2.59	2.54	2.15

The interaction of IQ × aids was also not significant as is shown in Table 49. The children in the low IQ range (78-100) only made a mean score of approximately 63 per cent on the verbal problems, or about 3.8 out of 6 problems. The children in the other two IQ ranges were not highly successful (mean scores of

Table 49

Interaction of IQ and Aids: Mean Scores

IQ	Aids		
	Physical	Pictorial	Verbal
1	2.64	2.56	2.32
2	2.71	2.65	2.26
3	2.43	2.41	1.88

below 80 per cent) on the verbal problems but were certainly more successful than those in the lowest IQ range.

Table 50 represents the interaction of levels × IQ × aids. This interaction is not statistically significant, but there are, however, interesting observations to be made. The children in Level 1, IQ range of 114-140 did just as well on the verbal problems as on the problems with accompanying aids. The children in Level 2 of the same IQ range also did well on the verbal problems. Due to the significance of the IQ factor and the lack of interaction between IQ and aids, the mean scores on the verbal problems generally steadily decrease across IQ within any given level. This is especially true for the first, second and fourth levels. The high IQ group in Level 3 performed much like the low IQ group in that same level which is attributable to an artifact of the sample. It is significant that the high IQ group of Level 4 performed no better than any IQ group of Levels 1, 2 and 3. However, the children at Level 1 in the low IQ group performed at least as well as the children at Levels 3 and 4 in the high IQ group, at least as well as the children in Levels 2, 3 and 4 in the middle IQ group, and, of course, no worse than the children in Levels 2, 3 and 4 of the low IQ group. Of the children at Level 4, there was not much difference between the first and second IQ groups, but the third IQ group warrants special discussion. These children had a mean score of only about 62 per cent for the problems with physical aids, 71 per cent for the problems with pictorial aids, and 39 per cent for the verbal problems. The curriculum these children have been through clearly has not given them a mastery of addition problems. Moreover, the first two IQ groups in Level 4 are also performing minimally. For them, the use of visual aids did not seem to be as helpful (relative to the scores on the verbal problems) as for the low IQ group. Among other groups that did not perform highly are those children at Level 2, IQ group 3; Level 3, IQ group 3; and perhaps Level 3, IQ group 1, which again can be attributed to an artifact of the sample.

TRANSFORMATION VS. NO TRANSFORMATION

In this section, results of the main effect of a transformation, described or actually present, versus no transformation, hereafter referred to as Factor D, will be given, thus continuing to subdivide the effect of the six tests. The reliabilities of these two subtests of .65

Table 50

Interaction of Levels, IQ and Aids: Mean Scores

Level	IQ	Physical	Pictorial	Verbal
1	1	2.59	2.59	2.64
	2	2.86	2.96	2.36
	3	2.77	2.73	2.14
2	1	2.91	2.91	2.54
	2	2.73	2.54	2.27
	3	2.59	2.32	2.04
3	1	2.59	2.54	2.09
	2	2.77	2.77	2.27
	3	2.50	2.46	2.14
4	1	2.46	2.18	2.00
	2	2.46	2.32	2.14
	3	1.86	2.14	1.18

Table 51

ANOVA for the Main Effect of Factor D

Source of variation	d.f.	MS	F
D	1	13.657	21.86***
D × Levels	3	.414	<1
D × IQ	2	.077	<1
D × Levels × IQ	6	.495	<1
D × Subj. w. groups	120	.625	

***p < .01, Conservative Test

Table 52

Interaction of Levels and Factor D: Mean Scores

Level	Factor D: Transformation	Factor D: No Transformation
1	2.75	2.51
2	2.74	2.34
3	2.58	2.34
4	2.17	1.99
Mean	2.56	2.30

and .81 are given in Table 27. The results of the interactions of D with levels and with IQ along with the interaction of D by levels and IQ will be also given. Table 51 gives the analysis of variance.

The main effect of Factor D is highly significant using the conservative test. Table 52 gives the means for this factor.

Due to the fact the children tested had been through an arithmetic curriculum which had given children problems which involved both a transformation and no transformation, it could be possible that Factor D would tend to become insignificant. The fact it remained significant should support the hypothesis that having a transformation, described or actually present, does, in fact, facilitate problem solving for the children. The question arises, then, does it facilitate problem solving more for some children than others? The answer for the children at the four levels studied is, statistically speaking, no, because of the insignificant interaction of levels and Factor D, shown in Table 52.

The interaction of IQ and Factor D was also insignificant so that the variable operated much the same for the children among the three IQ groups, as is shown in Table 53.

Table 54 shows the results of the insignificant interaction of levels, IQ and Factor D.

Table 53

Interaction of IQ and Factor D: Mean Scores

IQ	Factor D: Transformation	Factor D: No Transformation
1	2.65	2.36
2	2.65	2.42
3	2.37	2.11

Table 54

Interaction of Levels, IQ and Factor D: Mean Scores

Level	IQ	Transformation	No transformation
1	1	2.82	2.39
	2	2.76	2.70
	3	2.67	2.42
2	1	2.88	2.70
	2	2.79	2.24
	3	2.54	2.09
3	1	2.52	2.30
	2	2.73	2.49
	3	2.48	2.24
4	1	2.39	2.03
	2	2.33	2.27
	3	1.79	1.67

Due to the way the pretest was constructed, the children had to make a comparison of two states for every item of each subtest. When the children made a comparison of the two states in Item 4 of each subtest, they had to supply the transformation of either counting or setting up a one-to-one correspondence in order to make a correct response on the item, disregarding guessing. It has been observed that the children of Levels 1, 2, and 3 had very good probabilities of not guessing when responding correctly to at least one of the three last items of each subtest, while the children on Level 4 had a very good chance of not supplying the types of transformation described above. Moreover, it has been noted that curriculum builders work on the assumption that children think in terms of action (transformations), and develop at the outset the operation of addition in terms of action situations, and then progress into no action situations, with the heavier emphasis on the action oriented problems. One would hypothesize then, that a statistical relationship should occur as an interaction of levels × D, where the difference in the means of the problem with a transformation and those problems with no transformation are smaller for at least Level 1 than for Level 4. This same hypothesis could be formulated for IQ × D and IQ × levels × D. As has been noted, these interactions are all insignificant, which says that Factor D operates the same for all groups studied. This does not, of course, invalidate the assumption that children think in terms of action. However, the fact that there were 96 problems which involved a transformation in the workbook the children used and at most 52 which did not involve a transformation raises the question of whether the significance of Factor D is not the result of the relative emphasis placed on the two types of problems in the curriculum.

THE INTERACTIONS OF AIDS AND FACTOR D

In this section, the remaining effect of the six tests, the interaction of aids and Factor D, is given along with the possible interactions of levels or IQ by aids and Factor D, for which the analysis of variance is given in Table 55.

Table 55

ANOVA for the Interactions of Aids and Factor D

Source of variation	d.f.	MS	F
Aids × D	2	3.702	15.67***
Levels × Aids × D	6	.379	1.60
IQ × Aids × D	4	.083	< 1
Levels × IQ × Aids × D	12	.399	1.69
Aids × D × Subj. w. Groups	240	.236	

***$p < .01$, Conservative Test

The interaction of aids × D is highly significant; mean scores are given in Table 56. In view of the facts that some of the six subtests have low reliabilities associated with them, and that Problem 17 was much more difficult than either 16 or 18, the interaction

Table 56

Interaction of Aids × D: Mean Scores

D	Aids		
	Physical	Pictorial	Verbal
Transformation	2.71	2.56	2.41
No Transformation	2.48	2.52	1.89

must be interpreted with caution. It may not, in fact, exist with a more reliable subtest. This does not, however, invalidate the main effect of Factor D nor the main effect of aids since these subtests had substantial reliability coefficients associated with them.

The interaction of levels by aids and Factor D is given in Table 57. This interaction is not statistically significant as shown in Table 55, but there are observations that are of interest. The verbal problems with no transformation turned out to be considerably more difficult for the Level 1, 2, and 3 children than the corresponding verbal problems with a transformation, which may be due to the fact that in the verbal problems with a transformation, the two sets that corresponded to the addends in all three problems had equivalent objects in them and the joining objects were going to be doing the same thing as the objects at rest, which was an accurate reflection of the type of experience given to the children by the particular curriculum in which they participated. But in the verbal problems with no transformation there was the one relatively difficult problem in which there were two sets of equivalent objects (kittens) but the kittens were doing different things which could have placed them into the category of non-equivalent objects, so that the children had no equivalent situation in which to place both sets. It has been already discussed that the differences in the means of the verbal problems with no transformations and verbal problems with a transformation was perhaps over-emphasized by this one problem. We now see that the magnitude of over emphasis was approximately the same for all groups involved, which could indicate a training factor, as noted above. Also, the possibility of an over-emphasis may be partially explained on the basis of a quotation by Dodwell ". . . ability to answer correctly questions which involve simultaneous consideration of the whole class and its (two) component subclasses, appears to develop to a large extent independently of an understanding of the concept of cardinal number . . ."[109] If an over-emphasis exists, it could be the result of the greater emphasis of training on

Table 57

Interaction of Levels by Aids and Factor D: Mean Scores

Level	Transformation			No Transformation		
	Physical	Pictorial	Verbal	Physical	Pictorial	Verbal
1	2.85	2.82	2.58	2.64	2.70	2.18
2	2.85	2.70	2.67	2.64	2.48	1.91
3	2.79	2.61	2.33	2.46	2.58	2.00
4	2.33	2.12	2.06	2.18	2.30	1.48

addition situations involving a transformation using sets with equivalent objects or the possibility that the ability to simultaneously consider partial classes and the whole class develops independently of cardinal number. An investigation of the interaction of IQ × aids × D (Table 58) could perhaps throw more light on the problem. Since within any IQ group all levels are present, any observation made cannot be attributed to the effect of levels. This table is quite similar to Table 57 in that there is a larger difference between the means for the verbal problems with a transformation and the verbal problems without a transformation, than either physical or pictorial. This seems to give support to the fact that if the over-emphasis of the difficulty of the verbal problems with no transformation is in fact a true over-emphasis, then the training of the children is playing an important role in their performance on the problems. Because of the retardation of the children at Level 4 and those in IQ Range 3 when solving problems, Table 59, which gives a four way interaction, is of special interest, even though the interaction is insignificant. This table certainly does not support the fact that the visual aids helped the Level 1, high IQ children when solving problems. There are also other groups for which the visual aids did not facilitate problem solving. In the problems involving a transformation, children at Level 4, IQ groups 1 and 2 did about the same on all types of problems regardless of the aids used. The same can be said for the children at Level 2 in the IQ groups 1 and 2. However, the same observation does not hold for the children at Level 3 in the first two IQ groups, for any of the children in the case of the problems with no transformation, and for any of the children in IQ group 3, except for Level 3.

The children at Level 4, IQ group 3 did extremely poorly on the verbal problems with no transformation. They also did poorly on the verbal problems with a transformation and certainly did not score high on the problems with accompanying visual aids, regardless of whether the problems had a transformation or not. Children at Level 4 across all IQ groups generally did poorly on the verbal problems with no transformation. These observations are made with due regard to the fact that the interaction was not significant and that three of the six subtests discussed can be considered as fairly unreliable. However, they are observations which indicate considerable group fluctuation and thereby warrant further research.

THE PERFORMANCE OF THE CHILDREN IN THE FOUR LEVELS AND IN THE THREE IQ GROUPS ON THE TEST OF ADDITION FACTS

A two-way analysis of variance was performed to detect any possible differences in the mean performance of the children in the four levels of conservation of numerousness and in the three IQ groups. The ANOVA was considered as a fixed model so that the within group variation was used as the error term for each main effect as well as the interaction term. Table 60 gives the results of the ANOVA which shows that the mean performances of the children in the three IQ groups and the four levels on the test of addition facts were of sufficient difference (see Table 61) so as to be significant at the .05 level of significance. An inspection of the means (see Table 61) shows that the children in Level 4 had only approximately an average score of 76 per cent while the next lowest mean was approximately 89 per cent (Level 3 children). The interaction of IQ and levels was not significant, but due to the significance of both levels and IQ, it is to be expected that the mean performance of the children in Level 4, IQ group 3 should be considerably lower than that of many of the other children, which is the case as is shown in Table 61.

Table 58

Interaction of IQ by Aids and Factor D: Mean Scores

IQ	Transformation			No Transformation		
	Physical	Pictorial	Verbal	Physical	Pictorial	Verbal
1	2.80	2.57	2.59	2.48	2.54	2.04
2	2.77	2.66	2.52	2.64	2.64	2.00
3	2.54	2.46	2.11	2.32	2.36	1.64

Table 59

Interaction of Levels × IQ × Aids × D: Mean Scores

Level	IQ	Transformation			No Transformation		
		Physical	Pictorial	Verbal	Physical	Pictorial	Verbal
	1	2.82	2.64	3.00	2.36	2.55	2.27
1	2	2.91	2.91	2.45	2.82	3.00	2.27
	3	2.82	2.91	2.27	2.73	2.55	2.00
	1	3.00	2.82	2.82	2.82	3.00	2.27
2	2	2.73	2.02	2.82	2.73	2.27	1.73
	3	2.82	2.45	2.36	2.36	2.18	1.73
	1	2.82	2.64	2.09	2.36	2.45	2.09
3	2	2.91	2.82	2.45	2.64	2.73	2.09
	3	2.64	2.36	2.45	2.36	2.55	1.82
	1	2.55	2.18	2.45	2.36	2.18	2.55
4	2	2.55	2.09	2.36	2.36	2.55	1.91
	3	1.91	2.09	1.36	1.82	2.18	1.00

Table 60

ANOVA of Performance of Children in Four Levels and Three IQ Groups on a Test of Addition Facts

Source of variation	d.f.	MS	F
IQ	2	28.705	4.25*
Levels	3	18.081	2.68*
IQ × Levels	6	4.634	< 1
Within	120	6.756	
Total	131		

*p < .05

THE RELATIONSHIP OF THE SCORES ON THE ADDITION FACTS TEST AND THE PROBLEM SOLVING TEST

In order to gain more insight into the relationship between the scores on the problem solving test and the scores on the addition facts test, 24 correlation coefficients were computed, the first of which is the correlation of .49 (significant at the .01 level of significance) between the scores on the addition facts test and the total scores on the problem solving test for all 132 children.

In view of the fact that the mean performance of the children in the four levels and three IQ groups was significantly different both for the problem solving test and for the addition facts test, with those children in Level 4 and IQ group 3 having the lowest scores, it was decided to compute correlation coefficients

Table 61

Mean Scores on the Addition Fact Test by IQ and Levels

IQ	Level 1	2	3	4	Mean
1	10.00	9.55	8.64	8.73	9.23
2	9.73	9.55	9.55	7.82	9.16
3	7.44	8.73	8.64	6.36	7.80
Mean	9.06	9.27	8.94	7.64	

between total scores on the problem solving test and the number facts test for each of the four levels and the three IQ groups. The results of these computations are given in Table 62.

Coefficients for the children at Level 4 and for the children in IQ group 3 are substantial and indicate that for these children the scores on the two tests are not at all independent which in turn indicates that these children have not completely mastered the addition facts. Their difficulties with them can therefore be explained in a large part by their ability to solve addition problems in spite of the fact that the curriculum advocates that children must have automatic responses to certain basic facts.[110] These correlations are, however, consistent with the fact that children learn the addition facts as a result of solving many addition problems.

The correlation coefficients for the children at Levels 1 and 2 are small but significantly different ($p < .05$) from a zero correlation; however, the correlation coefficient for the children at Level 3 is not significant. These results are inconsistent and are not explainable in terms of the data collected. The insignificant correlation of .23 observed for those children in the top IQ group indicates that generally, for these children, the goal of learning the basic addition facts via problem solving has been achieved.

Further correlation coefficients were computed between the scores on the six problems without accompanying aids and the scores on the addition facts test and between the scores on the twelve problems with accompanying aids and the scores on the addition facts test. The results, given in Table 62, are quite surprising in that the correlations do not differ to any great extent. One would suspect that the correlation would be greater between the addition facts test and the problems with no accompanying aids than between the addition facts test and the problems with accompanying aids because the children did not need to know the addition facts to score a problem with accompanying aids correct. All they had to do was to count the total objects present. However, in the case of the problems with no accompanying aids, the children had greater need of the addition facts in order to obtain a correct score for the problem. They could, however, have counted mentally or on their fingers. Table 62 gives the analogous correlations for the children in the four levels and for the children in the three IQ groups. The only large difference in the corresponding correlations given in this table is for the children in the second IQ group. For these children, a knowledge of the addition facts apparently explained more of the variance of the scores on the problems with no aids than on the problems with aids. This does not necessarily mean, however, that more drill should be given on the addition facts for the children of this group. It certainly could be the case that more work should be given on the verbal

Table 62

Correlations Between the Scores on Six Problems Without Aids and the Scores on the Number Facts Test and Between the Twelve Problems with Accompanying Aids and the Scores on the Number Facts Test for the Children in the Four Levels and the Children in the Three IQ Groups

Problem Solving Test	Level 1	Level 2	Level 3	Level 4	IQ 1	IQ 2	IQ 3	Total
No Aids	.33	.39*	.16	.65**	.16	.47**	.52**	.46**
Aids	.31	.34	-.04	.56**	.24	.14	.54**	.41**
Total Test	.39*	.39*	.05	.68**	.23	.36*	.60**	

*p < .05
**p < .01

interpretation by the children of the pictorial or physical representation of the problem or more work on constructing problems on their own, both of which give added emphasis on the addition facts. Again, substantial correlations are present, both in the case of problems with accompanying aids and no accompanying aids, for the children at Level 4 and for the children in IQ group 3. These four correlation coefficients are no doubt lowered by the relative emphasis on drill on the addition facts.

VII

CONCLUSIONS

The results of this study have many implications both for actual classroom practice as it applies to the arithmetic curriculum of the elementary school and for more detailed research pertaining to the numerous unresolved problems in learning arithmetic. This duality of practice and research will be discussed as they pertain to the four levels of conservation of numerousness defined for the purpose of this study, the three IQ groups, and to each question asked in the statement of the problem.

THE FOUR LEVELS OF CONSERVATION OF NUMEROUSNESS

The extensive reliability study of the three tests of conservation of numerousness conducted and reported in Chapter V indicates that it should be possible to construct a test of conservation of numerousness with a higher internal consistency reliability coefficient. The three items of each test for which the children had to compare two sets of eight objects had very substantial inter-item correlation coefficients, significant beyond the .01 level of significance. Also, Item 3 of Test 1 (a comparison of 6 and 8 styrofoam balls, with the rectangle of 6 the larger of the two), Item 1 of Test 2 (a comparison of six and eight checkers in line segments of the same length), Item 3 of Test 2 (a comparison of six and eight checkers with the line segment of 6 the larger of the two), and Item 3 of Test 3 (a comparison of six and eight blocks in circles, with the circle of eight having the longer diameter of the two) all have significant inter-item correlations in pairs and also each of these items correlates significantly with the three items for which the children had to compare two sets of eight objects. Using only these seven items, an internal-consistency correlation coefficient of .74 is obtained. A test twice as long has an estimated internal—consistency reliability coefficient of .85. Moreover, it should be possible to construct more items which require comparison of two sets with the same number of objects in each set that correlate well with the above seven items. After the above research has been completed, an operational definition, much like that used in this study, of levels of conservation of numerousness must then be made and a test-retest reliability computed. It would then be desirable to attempt to construct a group test using analogous items for the sake of expediency of administering the test. It is felt that the above research is necessary in order to make the test of conservation of numerousness amenable for use in the elementary schools. It must be emphasized, however, that even though it is desirable to carry on further research on the test of conservation of numerousness, this in no way invalidates the results of the present study. Sufficient evidence has been given to support the internal-consistency reliability of the three tests (a coefficient of .69 was obtained) and the test-retest reliability of the four levels (A short time interval test-retest reliability of .78 was obtained and too, the test partitioned two independent samples of first graders in much the same way over a five month time interval.).

In view of the fact that the pretest did partition two independent samples in the same way over a five month time interval and in view of the substantial short time interval test-retest reliability obtained and with consideration given to the substantial longer time interval (3 months) test-retest reliability that Dodwell reports for his test, there is no reason to believe that a good test-retest reliability coefficient cannot be obtained with a revised version of the pretest (or for that matter, with the pretest used) over a period of six to eight months. Hence, when considered along with IQ scores from the Kuhlmann-Anderson IQ test, it appears that excellent prediction of relative success in solving addition problems and learning addition facts can be made for children entering the first grade. The prediction of the relative success of children in achieving other important outcomes of the first grade arithmetic curriculum (such as order of numbers, the ability to write a number sentence for a given problem, measurement, numeration system, etc.) all must await further research, but it now appears that the relative success of children in achieving these outcomes may be also predicted very well.

QUESTIONS ASKED IN THE STATEMENT OF THE PROBLEM

For convenience, each question asked in the statement of the problem that had a significant F ratio associated with it will be repeated and followed by a discussion in which implications for classroom practice or further research will be given, whenever they are relevant and appropriate.

QUESTION 1. Is the mean performances of children different for the six described arithmetic problems involving an additive structure?

The verbal problems with no transformation were significantly more difficult than all other problem types. Also, the problems with accompanying physical aids with a transformation were significantly easier than the problem types with accompanying physical aids with no transformation, accompanying pictorial aids with no transformation, and verbal problems with a transformation. At the present time, due to the low reliabilities reported, the above observations are in question. Further research with longer and thus more reliable tests is needed in order to completely affirm the above conclusions. However, if the above conclusions are valid (at present there is no reason to believe they are not), then they have implications for the type of research that should be undertaken to determine the optimal types of classroom experience first grade children should have. As noted, the arithmetic curriculum the children in the study participated in stressed problems with physical and pictorial aids with a transformation more than the same type of problems without a transformation. Also, there was not much stress (relatively speaking) placed on the solution of verbal problems. The above emphases perhaps are in part reflected by the significant differences obtained between the six problem types as the problems with accompanying pictorial aids with a transformation were neither statistically more difficult than the same kind of problems with physical aids nor statistically easier than the same kind of problems without accompanying aids while the problems with accompanying pictorial aids without a transformation were not statistically more difficult than the same kind of problems with physical aids, but were statistically easier than the same kind of problems without accompanying aids. This indicates that perhaps the greater emphasis placed on problems with a transformation in the curriculum may be having an equilizing effect on the ability of children to solve problems in varying degrees of abstraction. At this time it is not clear, based on the results of this study, that the same relative emphasis on problems without a transformation would not result in a better performance by the children on those problems and in particular, a better performance on the verbal problems without a transformation. Since more emphasis on the problems without a transformation would undoubtedly result in less emphasis on problems with a transformation, then research must also be conducted on an optimal balance of the two types of problems for different groups of children, which is discussed in the next few questions.

QUESTION 2. Is the mean performance of the children in each of the 12 groups different when solving arithmetic problems?

The mean performance of the children in the group defined by Level 4 and IQ group 3 scored significantly lower than all other groups except the four groups defined by Level 4, IQ groups 1 and 2; and IQ group 3, Levels 2 and 3. These latter differences were, however, very close to being significant. Since the reliability of the total problem solving test was substantial (.83), one can indeed safely conclude that even though all the children have gone through the same arithmetic curriculum, they have not all gained the same amount in terms of solving addition problems. (The highest scoring groups had a mean score of 93 per cent and the lowest scoring group a mean score of 58 per cent.) More detailed implications for research and classroom practice are contained in the discussion of the next question.

QUESTION 4. Is the mean performance of the children different in the four levels of conservation of numerousness?

The children in Level 4 performed significantly lower than the children in the top three levels. In terms of percentages, the mean performance of the children in Level 4 was 69 per cent and the mean performances of the children in Levels 1, 2 and 3 were 88, 85, and 82 per cent respectively. The differences in the last three percentages was not significant. The above observation supports quite well a comment made by Howard Fehr, quoted earlier that ". . . without a host of well developed concepts, it is very unlikely that a problem can be solved . . ."

Moreover, general intelligence also appears to be playing a vital role in childrens ability to solve problems as the mean performances of the children in the three IQ groups was significantly different, where the means in terms of percentages were 75, 85, and 83 per cent for the IQ groups 78-100, 101-113, and 114-140 respectively, where the mean of 75 per cent differed significantly from each of the other two, which answers Question 7: Is the mean performance of the children different in the three IQ groups? The interaction of the four levels and the three IQ groups was not significant which indicates that within any level, the children in the lowest IQ group may be expected to perform less well than the children in the two higher IQ groups. There are, then, at least 12 groups of first grade children with which curriculum builders must concern themselves when constructing an arithmetic curriculum. In view of the results, it certainly is not clear that the present first grade arithmetic curriculum is the optimal curriculum for each of these groups. It is entirely possible that a different type of curriculum is in order for the three groups of children in Level 4, which constitute 36 per cent of the sample, than for the other nine groups of children. Too, based on the results of this study, the curriculum for the children in the IQ range 78-100 in Level 4, which constitute 20 per cent of the sample, perhaps should be different than for the two other IQ groups in the same level. Moreover, it is entirely feasible those children in IQ group 78-100 who are in Levels 2 and 3 should be treated differently than those children in the two higher IQ groups across the same two levels. Hence, based on the results of this study, there is a total of three categories of children for which it can be justified that the types of experiences presently being provided produce different results:

1) The seven groups of children (52 per cent of the sample) that had mean scores above 80 per cent and whose mean scores were not statistically different. These seven groups are defined by Levels 1, 2, and 3 in IQ groups 101-113 and 114-140 and Level 1, IQ group 78-100. It must be emphasized, however, that there are considerable fluctuations within these seven groups which is apparent by a range in the means of 12 per cent (81 to 93 per cent). More will be said later about these seven groups and the groups in the two categories that follow.

2) The four groups of children (28 per cent of the sample) that had a mean score between 74 and 79 per cent defined by Level 4, IQ groups 101-113 and 114-140 and IQ group 78-100, Levels 2 and 3. This category denotes a "middle" category in that the mean performance of the children in the groups that constitute it were not statistically different from the mean performances of the seven groups of Category 1 nor statistically different from the mean performance of the children in the group of Category 3 below. However, the mean performance of the children in the group of Category 3 was statistically different from the mean performance of the children in seven groups in Category 1.

3) The one group of children (20 per cent of the sample) that had a mean score of 58 per cent defined by Level 4, IQ group 78-100.

QUESTION 5. Is the mean performance of children different for the problems involving the three levels of visual aids: 1) physical objects, 2) pictorial objects, and 3) no visual aids?

The problems with no accompanying aids were significantly more difficult than either of the other two types of problems for all children involved. The lack of an interaction of levels and aids; IQ and aids; and levels, IQ, and aids indicates that the variable of aids operated much the same way for all groups of children of interest for this study. There was, however, the exception that the children in Level 1, IQ group 114-140 performed much the same at all three levels of aids indicating these children at the time of year this study was conducted were able to work very effectively without the presence of aids. (This observation was made with due regard to an insignificant interaction but does warrant further research.) It may be the case, however, that these children are capable of working at a higher level of abstraction sooner than they are expected to in the present curriculum. That is, it is entirely possible that these children could be moved through an arithmetic curriculum at a higher level of abstraction without as much attention given to the visual aids as is presently being given, and thereby, be moved through the curriculum more rapidly. The same thing may be also true for the children at Level 2 in the IQ groups 114-140. Twenty-two per cent of the sample was in these two groups.

The children in Level 1, IQ groups 2 and 3, Level 2, IQ group 2, and Level 3, IQ groups 1 and 2 all have quite similar profiles with reference to the variable of aids. While it certainly is possible that many of these children are capable of being treated much like those children in the two groups just discussed, in general, the children in the above five groups did worse on the verbal problems than on the problems with aids. (Again, no significant interaction was present to justify the conclusion, but it does warrant further research.) Due to the insignificant differences observed among the performances of the children in all seven of the above groups, one cannot say the latter five groups should be treated a lot differently than the first two groups in the curriculum, but due to the fact of the relatively lower performance on the verbal problems, it certainly could be the case that the latter five groups need more experience with visual aids than the first two. However, it may be that too much emphasis is being placed on the visual aids, both physical and pictorial, for these children so that perhaps the same thing may be accomplished without as much work in the realm of the concrete. Thirty per cent of the sample was in these five groups. The seven groups just discussed are in Category 1 defined earlier.

The mean performances of those groups of children in Category 2—Level 2, IQ group 3; Level 3, IQ group 3; Level 4, IQ groups 1 and 2—did not, as has been noted, differ significantly from the mean performances of the seven groups in Category 1 nor from the one group in Category 3 so that the groups in Category 2 may be thought of as "middle" groups. In view of the mean scores the children in these groups obtained on the variables of aids, it is feasible that further work in the realm of concrete visual aids would be desirable. This is especially true of those children in Category 3, that is, those children in Level 4, IQ group 3, which constituted 20 per cent of the sample.

QUESTION 6. Is the mean performance of children different for the problems describing a transformation and the problems that do not describe a transformation?

The problems that involved a described transformation turned out to be significantly easier for the children than the problems that did not involve a described transformation. However, due to the lack of significant interactions, no conclusions could be drawn relative to the assumption made on the part of curriculum builders that children think in terms of action and thereby ought to have more problems involving a transformation than problems not involving a transformation. More research is needed in order to ascertain the optimal balance, if any, of problems with a described transformation and without a described transformation for the three categories of children defined earlier. Even though the interaction of levels, IQ, and Factor D was insignificant, there is a slight trend for the difference of the means of the groups of children in Categories 1 and 2 to be larger than the difference of the means of the group of children in Category 3.

QUESTION 12. Are the differences of the mean performance of the children the same for the problems describing a transformation across the three levels of visual aids?

The significance of this interaction remains in doubt due to three fairly unreliable subtests. However, if the subtests are in fact true representations of the relative ability of children to solve arithmetic addition problems, due to the insignificant higher-order interactions of aids by Factor D and levels or IQ, than it certainly seems advisable that experimentation be performed in order to define the optimal sequence of the combination of the two variables of aids and Factor D for each category of children discussed earlier.

QUESTIONS 16 AND 17. Is the mean performance of the children different on a test of number facts in the three IQ groups and the four levels of conservation of numerousness?

The mean performance of the children in IQ group 78-100 was lower than for each of the two other IQ groups (78, 92, and 92 per cent respectively). The differences of the mean performances of the children in the four levels was significant at the .051 level of significance. (The range of mean performances was 76 per cent to 93 per cent.) These results indicate that even though drill on the addition facts is emphasized, for many children, the facts are still not automatic. It is quite significant that the children in Level 4, IQ group 78–100 only had a mean score of 64 per cent, which indicates that these children are experiencing great learning difficulties not only with problem solving, but with learning the addition facts.

QUESTION 18. Is there a significant correlation between children's ability to solve addition problems and their knowledge of arithmetic facts ?

Twenty-four correlation coefficients were calculated in view of the above question. The correlation between total scores on the problem solving test and scores on the number facts test of .49 ($p < .01$) indicates a substantial relationship between the two tests. Since the curriculum in which the children were involved operates on the basis of obtaining the addition facts from problem solving, this relationship was expected. Since the curriculum also expects automatic responses from the children in the case of the addition facts, drill work was interspersed along with the discovery process which certainly could have resulted in a lower correlation coefficient than otherwise would be the case.

The correlation coefficient of .68 obtained for the children in Level 4 indicates that the drill these children received has not been highly beneficial for memorizing addition facts. It does indicate that for these children a strong relationship (perhaps stronger than the correlation coefficient indicates due to the retarding effect drill on the addition facts would have on the correlation) between problem solving abilities and a knowledge of addition facts exists which implies that less attention be spent on memorizing the facts and more attention be given to solving problems relevant to the interests of the children. The correlation coefficient of .60 obtained for those children in IQ group 3 results also in the same conclusion for these children as that just given for the children in Level 4. A correlation coefficient was computed for the 11 children in Level 4, IQ group 3 which turned out to be .83. This correlation coefficient is, however, quite unstable since it was computed on only 11 children. But in view of the two given for the children of Level 4 and IQ group 3, it indicates that drill procedures for these 11 children have been quite ineffective, and their knowledge of addition facts is dependent upon their problem solving abilities.

Small, but significant ($p < .05$) correlation coefficients were obtained for children in Levels 1 and 2 and IQ group 2 which indicates that the curriculum has been more effective in rendering these two variables independent for these three groups of children and, too, for the two additional groups defined by Level 3 and by IQ group 1, which had insignificant correlation coefficients associated with them.

A surprising result of this correlation study was that the correlation of the scores on the problems with accompanying aids and the scores on the addition facts test was not much different than the correlation obtained on the scores of the problems without accompanying aids and the scores on the addition facts test. This indicates that the relative difficulty of the verbal problems cannot be explained on the basis of a knowledge of the addition facts, which lends support to the conclusion that the presence of visual aids does in fact facilitate problem solving for first grade children. This same phenomenon was observed across all levels and IQ groups except for IQ group 2, which had a higher correlation in the case of no aids than in the case of aids (.47 versus .14). These over-all small differences in the correlations indicate that the teachers may not be taking full advantage of the visual aids when teaching problem solving; that is, they may not be building in the children the ability to solve any addition problems with any accompanying visual aids independently of the knowledge of addition facts, perhaps by a counting process. The question remains, then, is it possible to teach children to solve arithmetic addition problems so effectively, by an appropriate teaching method, that they will be able to solve any addition problem independently of their knowledge of the associated addition facts which are, at best, isolated bits of information that should obtain only after the appropriate problem solving abilities are developed ?

APPENDIX A
THE TEST OF ADDITION PROBLEMS

A. Physical Aids with a Transformation

1. There are four jacks in a pile and four more are put with them. Now how many jacks are in the pile?
2. There are three airplanes in a hanger and five more are put with them. Now how many airplanes are in the hanger?
3. There are four pegs in a pegboard and two more pegs are put with them. Now how many pegs are in the pegboard?

B. Physical Aids Without a Transformation

1. There are three cars in one parking lot and there are three cars in another parking lot. How many cars are in these parking lots?
2. There are five cookies on one plate and there are two cookies on another plate. How many cookies are on the plates?
3. There are six blocks in one pile and there are two blocks in another pile. How many blocks are there in the piles?

C. Pictorial Aids with a Transformation

1. There are five pails on the floor. Two more are put with them. Now how many pails are on the floor?
2. There are two fish eating seafood. Five more fish swim up to eat with them. Now how many fish are eating seafood?
3. There are two cakes on a table. Six more cakes are put with them. Now how many cakes are there on the table?

D. Pictorial Aids Without a Transformation

1. There are four houses on one side of a stream and two houses on the other side of the stream. How many houses are by the stream?
2. There are five balls in a pile and three balls in another pile. How many balls are there in the piles?
3. There are two candles on a table and three candles on another table. How many candles are there on the tables?

E. No Accompanying Aids with a Described Transformation

1. Three ducks are swimming on a pond and four ducks join them. Now how many ducks are swimming on the pond?
2. Two rabbits are playing in the garden. Four rabbits come to play with them. Now how many rabbits are playing in the garden?
3. Mike picked five apples and put them in a basket. He then picked two more apples and put them in the basket. Now how many apples are in the basket?

F. No Accompanying Aids Without a Described Transformation

1. John has three pennies in one hand and four pennies in his other hand. How many pennies does he have in his hands?
2. There are some kittens in the kitchen. Two kittens are drinking milk and five kittens are sleeping. How many kittens are in the kitchen?
3. In a zoo, there are three bears in one cage and five bears in another. How many bears are there in the cages?

APPENDIX B

FREQUENCY DISTRIBUTION OF SAMPLE BY IQ AND LEVELS

IQ	Level 1	Level 2	Level 3	Level 4	Total Frequency
78					
79	2			1	3
80		1		2	3
81					
82				3	3
83		1		5	6
84				2	2
85				3	3
86				4	4
87				4	4
88				1	1
89			1	6	7
90			1	2	3
91			2	2	4
92	1	1	1	5	8
93	1	1	2	5	9
94		2	3	5	10
95		1	1	2	4
96	3	1	2	2	8
97	1	1	1	5	8
98	1	1	2	1	5
99	1	1	2	1	8
100	1	1	1	8	11
101			2	3	5
102	1	2	2	3	8
103		1	3	2	6
104	1	0	3	2	6
105	1	1	1	1	4
106		1	2	1	4
107	2	2	2	5	11
108	1	4	3	5	13
109	1		6	3	10
110	1	2	2		5
111	1		2	5	8
112		2	5	2	9
113	2	3	1	5	11
114		1	5	4	10
115	5	6	3	2	16
116	3		2		5
117	3	2	2	2	9
118	1	2		2	5
119		3	3	1	7
120	1	2	1		4

IQ	Level 1	Level 2	Level 3	Level 4	Total Frequency
121	1	1	2	2	6
122	2	3	1	1	7
123	2	3			5
124	1	1	2		4
125	3	1	1	2	7
126		1	1	1	3
127		1	1	1	3
128		2	1		3
129	3				3
130	2	2	1	1	6
131	1		1	1	3
132		1	1		2
133	3	1	2		6
134					
135			1		1
136	4	1		1	6
137	1				1
138		1			1
139		1			1
140	2			1	3

NOTES

1. See Maurice L. Hartung, Henry Van Engen, F. Glenadine Gibb, James E. Stochl, Ray Walch and Lois Knowles, Seeing through Arithmetic 1 (Chicago: Scott, Foresman and Company, 1964); Robert Lee Morten, Merle Gray and Myron F. Rosskopf, Modern Arithmetic through Discovery, Book One (New Jersey: Silver Burdett Company, 1965); and School Mathematics Study Group, Mathematics for the Elementary School, Preliminary Edition, Book 1, Parts 1, 2 and 3 (Stanford University, 1963).
2. See Minnimath Reports, I, Number 1, p. 6; and Paul Rosenbloom, A paper written stating the theoretical background of the Minnemath program (University of Minnesota).
3. Felix Hausdorff, Set Theory (New York: Chelsa Publishing Company, 1962), p. 11.
4. Ibid., p. 8.
5. J. G. Kemeny, H. Mirkil, J. L. Snell, and G. L. Thompson, Finite Mathematical Structures (New Jersey: Prentice-Hall Inc., 1959), p. 51.
6. Howard Eves and Carroll Newsom, An Introduction to the Foundation and Fundamental Concepts of Mathematics (New York: Rinehart and Company, 1958), p. 228.
7. Ibid., p. 237.
8. Ibid., p. 236.
9. Henry Van Engen, Maurice L. Hartung and James E. Stochl, Foundations of Elementary School Arithmetic (Chicago: Scott, Foresman and Company, 1965), p. 27.
10. Ibid., Chapter 2.
11. Ibid. Chapter 4.
12. Ibid., p. 56.
13. Howard Fehr, "The Teaching of Mathematics in the Elementary School" (A paper delivered at the Conference on the Analysis of Conceptual Learning, Research and Development Center for Learning and Re-Education, University of Wisconsin), October 18-21, 1965, pp. 19-20.
14. Fehr, p. 21.
15. Donald E. Ship, et al., Developing Arithmetic Skills (New Jersey: Prentice Hall, 1964), pp. 85-89.
16. Morten, et al.
17. Morten, et al., "Teachers Edition," 1962, pp. 13-16.
18. Jean Piaget, The Child's Conception of Number (London: Routledge and Kegan Paul, 1952), p. 167.
19. "Piaget Rediscovered," A Report of the Conference on Cognitive Studies and Curriculum Redevelopment, eds., Richard E. Ripple and Verne E. Rockcastle (Cornell University, School of Education, March, 1964), pp. 9-10.
20. Ibid., p. 8.
21. Ibid., p. 9.
22. Piaget, pp. 3-4.
23. Jan Smedslund, "Concrete Reasoning: A Study of Intellectual Development," Monographs of the Society for Research in Child Development, XXIX (Serial No. 93, 1964), 28.
24. Ripple and Rockcastle, p. 9.
25. Piaget, pp. 5-13.
26. Ibid., pp. 196-197.
27. Ripple and Rockcastle, p. 10.
28. Arthur J. Coxford, Jr., "Piaget: Number and Measurement," Arithmetic Teacher, X (Nov., 1963), 420.
29. J. Houston Banks, Learning and Teaching Arithmetic (Boston: Allyn and Bacon, 1959), p. 364.
30. Fehr, pp. 22-23.
31. Ibid., p. 23.
32. Henry Van Engen, "Twentieth Century Mathematics for the Elementary School." Arithmetic Teacher, VI (March, 1959), 71-76.
33. David Elkind, "The Development of Quantitative Thinking: A Systematic Replication of Piaget's Studies," The Journal of Genetic Psychology, XCVIII (1961), 37-46.
34. Piaget, p. 5.
35. Elkind, p. 37.

36. Piaget, p. 244.
37. _Ibid._, p. 5.
38. _Ibid._, p. 244.
39. _Ibid._, p. 5.
40. _Ibid._, p. 5.
41. P. C. Dodwell, "Children's Understanding of Number and Related Concepts," _Canadian Journal of Psychology_, XIV (1960), 191-195.
42. H. Van Engen and L. Steffe, _First Grade Children's Concept of Addition of Natural Numbers._ Technical Report No. 5, Research and Development Center for Learning and Re-Education, C-03, PE 5-10-154 (Madison: University of Wisconsin, 1966).
43. David Elkind, "The Development of the Additive Composition of Classes in the Child: Piaget Replication Study III," _The Journal of Genetic Psychology_, XVI (1961), 152-159.
44. _Ibid._, p. 53.
45. P. C. Dodwell, "Relations Between the Understanding of the Logic of Classes and of Cardinal Numbers in Children," _Canadian Journal of Psychology_, XVI (1962), 152-159.
46. _Ibid._, p. 158.
47. Dodwell, 1960.
48. P. C. Dodwell, "Children's Understanding of Number Concepts: Characteristics of an Individual and of a Group Test," _Canadian Journal of Psychology_, XV (1961), 30.
49. _Ibid._, p. 32.
50. Van Engen and Steffe, p. 5.
51. _Ibid._, p. 5.
52. Dodwell, "Children's Understanding...," pp. 30-31.
53. _Ibid._, p. 31.
54. Ripple and Rockcastle, p. 21.
55. _Ibid._, pp. 11-12.
56. Van Engen and Steffe, p. 9.
57. _Ibid._, p. 9.
58. Hartung, et al., pp. 317-324.
59. Piaget, _The Child's Conception of Number_, pp. 170-171.
60. Ripple and Rockcastle, p. 9.
61. Smedslund, "Concrete Reasoning," p. 32.
62. Marilyn Zweng, "A Study of the Performance of Second Grade Children on Four Kinds of Division Problems" (Unpublished doctoral Dissertation, University of Wisconsin), p. 71
63. Dodwell, 1961, p. 35.
64. Millie Almy et al., _Young Children's Thinking: Studies of Some Aspects of Piaget's Theory_ (Teachers College Press, Teachers College, Columbia University, New York, 1966), pp. 51-52.
65. Eileen M. Churchill, "The Number Concepts of the Young Child: Part 2," (Researches and Studies, Leeds University), 1958.
66. Elkind, 1961, pp. 39-40.
67. K. D. Feigenbaum, "Task Complexity and IQ as variables in Piaget's problem of conservation," _Child Development_, XXXIV (1963), 423-432.
68. Smedslund, p. 12.
69. Joachim F. Wohlwill, "A Study of the Development of the Number Concept by Scalogram Analysis," _The Journal of Genetic Psychology_, XCVII (1960), 352.
70. E. A. Lunzer, _Recent Studies in Britain Based on the Work of Jean Piaget_ (London: National Foundation for Educational Research in England and Wales, Occasional Paper No. 4), 1960, p. 33.
71. Van Engen and Steffe, p. 6.
72. Elkind, Replication Study I.
73. Smedslund, p. 12.
74. Herbert J. Klausmeier, _Learning and Human Abilities: Educational Psychology_ (New York: Harper and Brothers, 1961), p. 404.
75. _Ibid._, p. 407.
76. _Ibid._, p. 378.
77. George F. Thompson, et al., _Educational Psychology_ (New York: Appleton-Century-Crofts, Inc., 1959), p. 124.
78. _Mathematics 1-6 Curriculum Guides_ (Division of Instructional Services, Racine Public Schools, Racine, Wisconsin), p. A-4.
79. _Ibid._, pp. A-7, A-8.
80. Leslie P. Steffe, "The Performance of First Grade Children in Four Levels of Conservation of Numerousness and Three IQ Groups When Solving Arithmetic Addition Problems" (Unpublished doctoral dissertation, University of Wisconsin), pp. 51-57.
81. B. J. Winer, _Statistical Principles in Experimental Design_ (New York: McGraw Hill Book Company, 1962), Chapter 7.
82. _Ibid._, p. 350.
83. Samuel W. Greenhouse and Seymour Geisser, "On Methods in the Analysis of Profile Data," _Psychometrika_, XXIV (June, 1959), 95-112.
84. _Ibid._, p. 100.
85. _Ibid._, p. 102.
86. Julian C. Stanley, _Measurement in Today's Schools_ (New Jersey: Prentice-Hall Inc., 1964), p. 155.

87. Ibid., p. 156.
88. Robert M. Schoeht, "Unified Means, Standard Deviations and Correlation Program." (G2-Wisc STAT I), Feb. 1, 1964.
89. Ibid., p. 156.
90. Lee J. Cronbach, Essentials of Psychological Testing (New York: Harper and Row, 1960), p. 131.
91. Frank B. Baker, "An Intersection of Test Score Interpretation and Item Analysis," Journal of Educational Measurement, I (June 64), 24-25.
92. Ibid., p. 25.
93. Frank B. Baker, Empirical Determination of Sampling Distributions of Item Discrimination Indices and a Reliability Coefficient (Department of Educational Psychology, School of Education, University of Wisconsin, Contract OE-2-10-071, Nov. 62), pp. 11-12.
94. Baker, p. 25.
95. Baker, Empirical Determination of. . ., p. 29.
96. Ibid., p. 6.
97. Frank B. Baker, Test Analysis Package: A Program for the CDC 1604 - 3600 Computers (Laboratory of Experimental Design, Dept. of Educational Psychology, U. of Wis., June 1966), p. 6.
98. Ibid., p. 5.
99. Baker, Empirical Determination of. . ., p. 87.
100. Ibid., p. 87.
101. Baker, Test Analysis Package: . . ., June 66.
102. Ernest W. Tiegs and Willis W. Clark, Manual: California Achievement Tests, Lower Primary, (Los Angeles: California Test Bureau, 1957), p. 8.
103. Steffe, pp. 91-117.
104. Winer, p. 309-310.
105. Ibid., p. 83.
106. Cronbach, p. 218.
107. Ibid., pp. 218-220.
108. Ibid., p. 182.
109. Dodwell, 1962, p. 158.
110. Curriculum Guides, p. 3.

BIBLIOGRAPHY

Almy, Millie, et al. Young Children's Thinking: Studies of Some Aspects of Piaget's Theory. Teachers College Press, Teachers College, Columbia University, New York 1966.

Baker, Frank B. An Intersection of Test Score Interpretation and Item Analysis. Journal of Educational Measurement. 1964, 1 23-28.

Baker, Frank B. Empirical Determination of Sampling Distribution of Item Discrimination Indices and a Reliability Coefficient. Department of Educational Psychology, School of Education, University of Wisconsin, Contract OE-2-10-071, Nov. 1962.

Baker, Frank B. Test Analysis Package: A Program for the CDC 1604-3600 Computers. Laboratory of Experimental Design, Department of Educational Psychology, University of Wisconsin, 1966.

Banks, J. Houston. Learning and Teaching Arithmetic. Boston: Allyn and Bacon, 1959.

Churchill, Eileen M. The Number Concepts of the Young Child: Part 2. Researches and Studies, Leeds University, 1958 (b).

Coxford, Arthur J. Jr. Piaget: Number and Measurement. Arithmetic Teacher, Nov., 1963, 10, 419-427.

Cronbach, Lee Jr. Essentials of Psychological Testing. New York: Harper and Row, 1960.

Dodwell, P. C. Children's Understanding of Number and Related Concepts. Canadian Journal of Psychology, 1960, 14, 191-205.

Dodwell, P. C. Children's Understanding of Number Concepts: Characteristics of an Individual and of a Group Test. Canadian Journal of Psychology, 1961, 15, 29-36.

Dodwell, P. C. Relations Between the Understanding of the Logic of Classes and of Cardinal Numbers in Children. Canadian Journal of Psychology, 1962, 16, 152-160.

Elkind, David. The Development of Quantitative Thinking: A Systematic Replication of Piaget's Studies. The Journal of Genetic Psychology, 1961, 98, 36-46.

Elkind, David. The Development of the Additive Composition of Classes in the Child: Piaget Replication Study III. The Journal of Genetic Psychology, 1961, 15, 51-57.

Eves, Howard and Newsom, Carroll. An Introduction to the Foundation and Fundamental Concepts of Mathematics. New York: Rinehart and Company, 1958.

Fehr, Howard. The Teaching of Mathematics in the Elementary School. A paper delivered at the Conference on the Analysis of Conceptual Learning, Research and Development Center for Learning and Re-Education. University of Wisconsin, October 18-21, 1965.

Feigenbaum, K. D. Task Complexity and IQ as Variables in Piaget's Problem of Conservation. Child Development, 1963, 34, 423-432.

Feller, William. An Introduction to Probability Theory and Its Applications, New York: John Wiley and Sons, Inc., 1957.

Ferguson, George A. Statistical Analysis in Psychology and Education. New York: McGraw Hill Book Company, 1959.

Goodwin, William L. The Effects on Achievement Test Results of Varying Conditions of Experimental Atmosphere, Notice of Test, Administration, and Test Scoring. Technical Report No. 2. Research and Development Center for Learning and Re-Education, C-03, OE5-10-154. Madison: University of Wisconsin, 1965.

Greenhouse, Samuel W. and Geisser, Seymour. On Methods in the Analysis of Profile Data. Psychometrika, 1959, 24, 95-112.

Hartung, Maurice L., Van Engen, Henry, Gibb, Glenadine, Stochl, James E., Walch, Ray and Knowles, Lois. Seeing Through Arithmetic I. Chicago: Scott Foresman and Company, 1964.

Hausdorff, Felix. Set Theory. New York: Chelsa Publishing Company, 1962.

Kemeny, J. G., Mirkil, H., Snell, J. L. and Thompson, G. L. Finite Mathematical

Structures. New Jersey: Prentice-Hall Inc., 1959.

Klausmeier, Herbert J. Learning and Human Abilities: Educational Psychology. New York: Harper and Brothers, 1961.

Lowell, K. Growth of Basic Mathematical and Scientific Concepts in Children. London: University of London Press, 1961.

Mathematics 1-6 Curriculum Guides. Division of Instructional Services, Racine Public Schools, Racine, Wisconsin.

Minnimath Reports. University of Minnesota, Volume 1, Number 1.

Morten, Robert Lee, Gray, Merle and Rosskopf, Myron F. Modern Arithmetic Through Discovery, Book I. New Jersey: Silver Burdett Company, 1965.

Piaget, Jean. The Psychology of Intelligence, London: Routledge and Kegan Paul, Fourth Impression, 1964.

Piaget Rediscovered. A Report of the Conference on Cognitive Studies and Curriculum Redevelopment, eds. Ripple, Richard E. and Rockcastle, Verne E. Cornell University, School of Education, March 1964.

Rosenbloom, Paul. A paper written stating the theoretical background of the M·anemast program. University of Minnesota.

Schoeht, Robert M. Unified Means, Standard Deviation and Correlation Program G2-Wisc. STAT, Feb. 1, 1964.

School Mathematics Study Group. Mathematics for the Elementary School, Preliminary Edition, Book 1, Parts 1, 2 and 3. Stanford University, 1963.

Shipp, Donald E., et al. Developing Arithmetic Skills. New Jersey: Prentice Hall, 1964.

Smedslund, Jan. Concrete Reasoning: A Study of Intellectual Development. Monographs of the Society for Research in Child Development, Serial No. 93, 1964, 29.

Stanley, Julian C. Measurement in Today's Schools. New Jersey: Prentice-Hall, Inc., 1964.

Thompson, George F., et al. Educational Psychology. New York: Appleton-Century Crofts, Inc., 1959.

Tiegs, Ernest W. and Clark, Willis W. Manual: California Achievement Tests, Lower Primary. Los Angeles: California Test Bureau, 1957.

Van Engen, Henry. Twentieth Century Mathematics for the Elementary School. Arithmetic Teacher, 1959, 6, 71-76.

Van Engen, Henry, Hartung, Maurice L. and Stoehl, James E. Foundations of Elementary School Arithmetic. Chicago: Scott Foresman and Company, 1965.

Van Engen, H. and Steffe, L. First Grade Children's Concept of Addition of Natural Numbers. Technical Report No. 5, Research and Development Center for Learning and Re-Education, C-03, OE 5-10-154. Madison: University of Wisconsin, 1966.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company, 1962.

Wohlwill, Joachim F. A Study of the Development of the Number Concept by Scalogram Analysis. The Journal of Genetic Psychology, 1960, 97, 347-377.

Wohlwill, J. F. and Lowe, R. C. Experimental Analysis of the Development of the Conservation of Number, Child Development, 1962, 33, 153-169.

Zweng, Marilyn. A Study of the Performance of Second Grade Children on Four Kinds of Division Problems. Unpublished Ph. D. Dissertation, University of Wisconsin.